فراخوان راهکارهای تحول دیجیتال صنعت بیمه | فراخوان ۵

بخش فنی پروپوزال

معماری، مدل داده، حریم خصوصی، موتور کشف تقلب، امتیازدهی و برنامه اثبات در ۱۲ هفته

فراخوان ۵: کشف تقلب و تخلف در بیمه خودرو

سند مرتبط	پروپوزال کامل — بخش محصول و کسب‌وکار
تهیه‌کننده	**نام شرکت / تیم**
سرپرست فنی و داده	**نام و سمت**
تاریخ	آبان ۱۴۰۵
نسخه	۱,۰
طبقه‌بندی	محرمانه — مخصوص ارزیابی فنی

این سند شامل بخش محصولی–بیزینسی و بخش فنی است. نسخه ۱,۰ — آبان ۱۴۰۵

راهنمای این سند

این سند، بخش فنی پروپوزال فراخوان ۵ است و مستقل از بخش محصولی–بیزینسی قابل خواندن است. هدف آن، پاسخ به سه پرسش ارزیاب فنی است: راهکار چگونه ساخته می‌شود؟ چگونه از داده و حریم خصوصی محافظت می‌کند؟ و چگونه در ۱۲ هفته اثبات می‌شود؟

- **فصل‌های ۷ تا ۱۰:** معماری، مدل داده، حریم خصوصی و یکپارچه‌سازی.
- **فصل‌های ۱۱ تا ۱۳:** موتور کشف تقلب، امتیازدهی و اعتبارسنجی پایلوت.
- **فصل‌های ۱۴ تا ۱۷:** امنیت، اجرا، تیم، ریسک و نتیجه‌گیری.
- **پیوست‌ها:** نمونه API، کاتالوگ قواعد، چک‌لیست داده و فرم ارزیابی.

> **نکته برای ارزیاب**
>
> این سند عامدانه از ادعاهای غیرقابل اثبات پرهیز کرده است: گواهینامه‌ای که دریافت نشده، دقتی که اندازه‌گیری نشده، و داده‌ای که در اختیار نداریم. هرجا عددی آمده، فرض یا هدف مشخص است.

فهرست مطالب

راهنمای این سند 2

۷ معماری مرجع راهکار 5

۸ مدل داده و شناسه‌های مشترک 10

۹ موتور بی‌نام‌سازی و حریم خصوصی 15

۱۰ یکپارچه‌سازی و اتصال به سیستم‌های بیمه 19

۱۱ موتور کشف تقلب: الگوها، گراف و متن 22

۱۲ امتیازدهی، اولویت‌بندی و جریان کار 28

۱۳ اعتبارسنجی، آزمون و طراحی پایلوت 32

۱۴ امنیت، انطباق و حاکمیت داده 35

۱۵ برنامه اجرا و تیم 39

۱۶ ریسک‌ها، شاخص‌ها و ارزش اثباتی 43

۱۷ نتیجه‌گیری و درخواست 45

پیوست‌ها 47

بخش ب

راهکار فنی

معماری، داده، حریم خصوصی، موتور تشخیص، امتیازدهی و برنامه اثبات در ۱۲ هفته.

— معماری مرجع و مرز اعتماد
— مدل داده و شناسه‌های مشترک
— موتور بی‌نام‌سازی و حریم خصوصی
— یکپارچه‌سازی و API
— موتور کشف تقلب و گراف
— امتیازدهی، میز کار و اعتبارسنجی
— امنیت، اجرا، ریسک و نتیجه‌گیری

۷ معماری مرجع راهکار

معماری این راهکار بر یک اصل ساده استوار است: داده خام هویتی از بیمه‌گر خارج نمی‌شود، اما شواهد کافی برای تصمیم به شبکه می‌رسد. این فصل، لایه‌ها، مرز اعتماد، توپولوژی استقرار، فناوری و مشخصات غیروظیفه‌شناسی را مشخص می‌کند.

۷.۱ اصول معماری

جدول ۷-۱ — اصول و تصمیم‌های معماری

اصل	تصمیم طراحی	فایده
حریم خصوصی از طراحی	توکن‌سازی در نقطه ورود، عدم ذخیره شناسه خام، حداقل‌سازی فیلد	کاهش ریسک حقوقی و نشت
ماژولار	شش ماژول مستقل با قرارداد API	توسعه تدریجی و جایگزینی اجزا
first-API	همه قابلیت‌ها از طریق API قابل فراخوانی	یکپارچه‌سازی با سیستم بیمه‌گر
نقطه‌زمانی	ویژگی‌ها فقط از داده پیش از تاریخ اعلام خسارت	جلوگیری از نشت آینده و سوگیری
توضیح‌پذیر	هر امتیاز به قاعده، شاهد و نسخه متصل است	اعتماد کارشناس و ناظر
استقرارپذیر	prem-on، ابر خصوصی یا ترکیبی	انطباق با الزامات اقامت داده
مبتنی بر استاندارد	SQL ،OAuth2، OpenAPI ،JSON/REST	کاهش وابستگی به فروشنده

۷.۲ نمای کلی لایه‌ها

معماری از شش لایه و یک ستون حاکمیت تشکیل شده است. لایه یک، سیستم‌های منبع بیمه‌گران است. لایه دو، اتصال و اعتبارسنجی را انجام می‌دهد. لایه سه، مرز اعتماد و توکن‌سازی است. لایه چهار، داده کانونیکال، گراف و مخزن ویژگی را نگه می‌دارد. لایه پنج، موتورهای تشخیص را اجرا می‌کند. لایه شش، نتیجه را به میز کار، API و داشبورد می‌رساند. حاکمیت و ممیزی، عمودی بر همه لایه‌ها اعمال می‌شود.



شکل ۷-۱ — معماری مرجع شش‌لایه با مرز اعتماد و ستون حاکمیت

جدول ۷-۲ — لایه‌ها و مسئولیت‌ها

لایه	اجزای اصلی	مسئولیت	کنترل کلیدی
۱. منابع	بیمه‌نامه، خسارت، کارشناسی، پرداخت، CRM، اسناد	تولید داده خام	دسترسی خواندنی محدود
۲. اتصال	API Gateway، SFTP امن، کانکتور، اعتبارسنجی طرح	انتقال و اعتبارسنجی	mTLS، PGP، schema validation
۳. اعتماد	توکن‌سازی، KMS/HSM، هویت، ثبت هدف، DLP	حذف هویت و کنترل دسترسی	تفکیک وظایف، نرخ استعلام
۴. داده	انبار کانونیکال، گراف، مخزن ویژگی، دریاچه ممیزی	ذخیره حداقلی و قابل ردیابی	رمزنگاری، نسخه‌بندی
۵. تحلیل	قواعد، گراف، NLP، ناهنجاری، امتیاز	تولید شاهد و امتیاز	نسخه قواعد، آزمون‌پذیری
۶. مصرف	میز کار، API، داشبورد، کنسول ناظر	تحویل نتیجه و بازخورد	RBAC، حداقل افشا
حاکمیت	ممیزی، SIEM، پایش نشت، کیفیت داده	انطباق و پایداری	لاگ تغییرناپذیر، ممیزی سالانه

۷.۳ مرز اعتماد و تفکیک وظایف

کلید امنیت این طرح، «تفکیک سه نقش» است: **نگهدارنده کلید**، **اپراتور داده** و **تحلیل‌گر**. هیچ نقشی به‌تنهایی نمی‌تواند هویت یک فرد را به داده تحلیلی متصل کند. نگهدارنده کلید در نهاد امین/HSM است و به داده دسترسی ندارد؛ اپراتور داده، رکوردهای توکن‌شده را مدیریت می‌کند و کلید ندارد؛ تحلیل‌گر، الگوها را می‌بیند و نه هویت و نه کلید را در اختیار ندارد.

جدول ۷-۳ — نواحی اعتماد و کنترل دسترسی

ناحیه	داده	چه کسی دسترسی دارد	کنترل
ناحیه بیمه‌گر	داده خام کامل	کارکنان مجاز بیمه‌گر	DLP، سیاست داخلی، دسترسی نقش‌محور
ناحیه اعتماد	شناسه خام در لحظه توکن‌سازی	سرویس توکن و نگهدارنده کلید	HSM، لاگ، نرخ محدود، عدم ذخیره
ناحیه تحلیل	رکوردهای توکن‌شده و ویژگی‌ها	تحلیل‌گر و موتورها	بدون شناسه، حداقل‌سازی، ماسک
ناحیه مصرف	امتیاز، برچسب، کد دلیل	کارشناس، مدیر، ناظر	RBAC، حداقل افشا، ممیزی



شکل ۷-۲ — مسیر کنترل‌شده استعلام و پاسخ؛ هیچ داده خامی بازنمی‌گردد

۷.۴ توپولوژی استقرار

جدول ۷-۴ — گزینه‌های استقرار

گزینه	مزیت	محدودیت	کاربرد
هاب مرکزی شبه‌بی‌نام	ساده‌ترین برای پایلوت، تحلیل بین‌شرکتی کامل	نیاز به نهاد امین و قرارداد	پایلوت و تولید اولیه
استقرار در هر بیمه‌گر	داده از سازمان خارج نمی‌شود	تحلیل بین‌شرکتی ممکن نیست	تحلیل درون‌شرکتی
ترکیبی (هاب + کانکتور)	تعادل حریم خصوصی و تحلیل	پیچیدگی عملیاتی متوسط	پیشنهاد تولید
فدرال / رمزنگاری هم‌ریخت	حداکثر حریم خصوصی	پیچیدگی و هزینه بالا، بلوغ پایین	نقشه راه آینده

پیشنهاد: پایلوت در **هاب مرکزی شبه‌بی‌نام** داخل آزمایشگاه داده انجام شود؛ تولید به **مدل ترکیبی** حرکت کند: کانکتور سبک در بیمه‌گر، هاب تحلیلی در نهاد امین، و کنترل کلید در HSM مستقل. این مسیر، از یک‌جا رفتن به مدل فدرال پیچیده جلوگیری می‌کند و در عین حال مسیر ارتقا را باز نگه می‌دارد.

۷.۵ انتخاب فناوری

جدول ۷-۵ — انتخاب فناوری

لایه	پیشنهاد	جایگزین	دلیل انتخاب
دریافت داده	dbt + Airflow + Python	NiFi ،Spark	بالغ، متن‌باز، نیروی متخصص فراوان
ذخیره	MinIO + PostgreSQL	Oracle ،SQL Server	هزینه پایین، پشتیبانی داخلی، استاندارد
گراف	Apache AGE یا Neo4j	(پایلوت) NetworkX	پرس‌وجوی مسیر و خوشه؛ AGE روی PostgreSQL
جست‌وجو	OpenSearch	Elasticsearch	متن‌باز، جست‌وجوی متن فارسی
پردازش	Spark/Polars + Python	—	سرعت توسعه، مقیاس‌پذیری
NLP	ParsBERT + Hazm	مدل‌های عمومی	زبان فارسی، قابلیت تنظیم دقیق
قواعد	Python ارزیاب + YAML/DSL JSON	Drools	شفافیت، نسخه‌بندی، آزمون‌پذیری
سرویس	OPA + Keycloak + FastAPI	Spring Boot	سبک، مستند OpenAPI، سیاست‌محور
رابط کاربری	js.Next/React	Vue	اکوسیستم، سرعت توسعه
داشبورد	Superset / Metabase	Tableau	متن‌باز، استقرار داخلی
امنیت	Trivy ،ELK ،Wazuh ،HSM/Vault	—	مدیریت کلید، اسکن آسیب‌پذیری

۷.۶ مشخصات غیروظیفه‌شناسی

جدول ۷-۶ — مشخصات غیروظیفه‌شناسی

شاخص	هدف پیشنهادی	نحوه سنجش
زمان پاسخ استعلام	p95 < ۲ ثانیه	آزمون بار با ۱۰۰ درخواست همزمان
پردازش دسته‌ای	۱ میلیون رکورد در پنجره شبانه	آزمون حجم و پنجره زمانی
دسترس‌پذیری	≥ ۹۹٫۵٪ در پایلوت	پایش ماهانه
RPO / RTO	۴ ساعت / ۱ ساعت	آزمون بازیابی
مقیاس‌پذیری	۱۰ میلیون پرونده، ۵۰ بیمه‌گر	طراحی افقی و آزمون شبیه‌سازی
نگهداری داده	۳۶ ماه تحلیل، حذف پس از دوره	سیاست نگهداری و آزمون حذف
پشتیبان‌گیری	روزانه، رمزنگاری‌شده، خارج از سایت	آزمون بازیابی فصلی

۷.۷ چرا این معماری در ۱۲ هفته قابل اجراست

- **بدون جایگزینی سیستم فعلی:** اتصال با batch و API انجام می‌شود؛ بیمه‌گر سیستم خود را نگه می‌دارد.
- **شروع با حداقل داده:** مدل کانونیکال فقط فیلدهای لازم برای چهار الگو را می‌خواهد؛ نه کل پایگاه داده.
- **موتور بی‌نام‌سازی موجود:** کد مهندسی فعلی، نقطه شروع لایه توکن‌سازی است و از صفر ساخته نمی‌شود.
- **تحلیل روی داده تاریخی:** پایلوت سایه نیازی به تغییر فرآیند عملیاتی در هفته اول ندارد.

- **معماری ماژولار:** اگر NLP یا گراف در پایلوت تأخیر داشته باشد، امتیازدهی پایه و میز کار همچنان قابل ارائه است.

۸ مدل داده و شناسه‌های مشترک

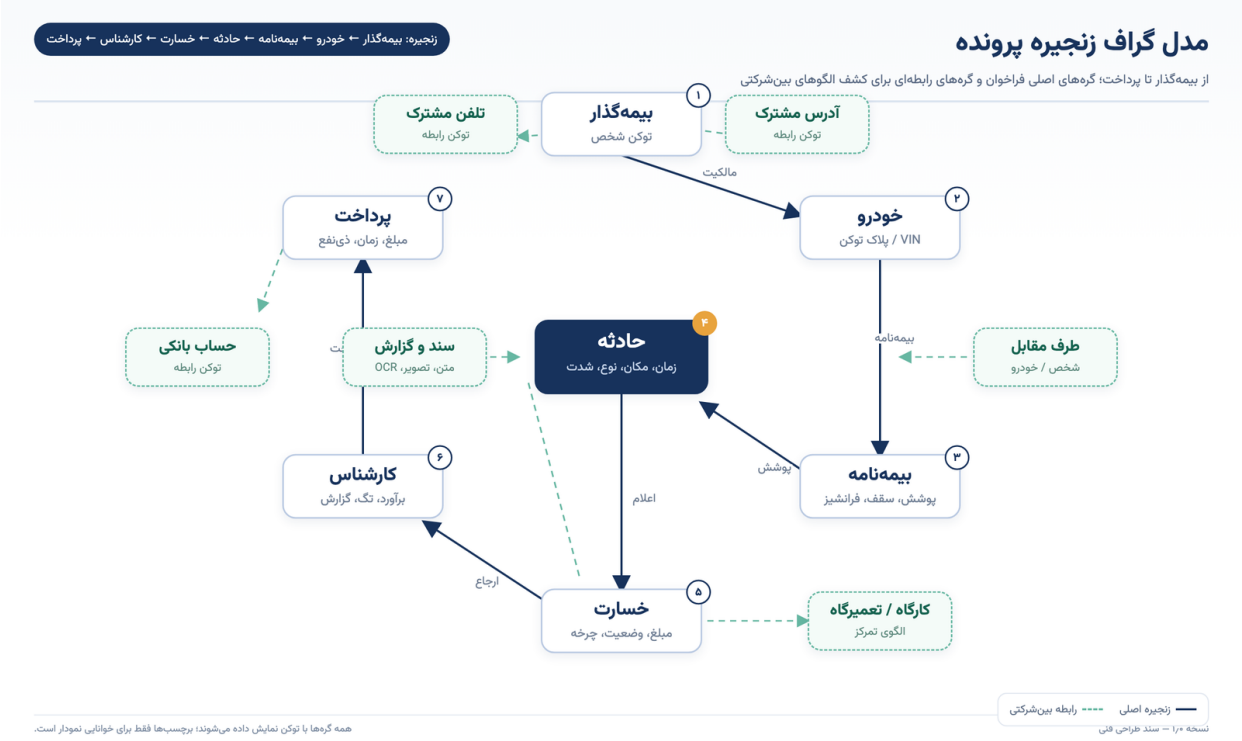
بدون یک مدل داده مشترک، هر بیمه‌گر داده را به زبان خودش می‌فرستد و تحلیل بین‌شرکتی غیرممکن می‌شود. این فصل، موجودیت‌های کانونیکال، زنجیره هفت‌گانه فراخوان، توکن‌ها، تطبیق رخداد، استانداردسازی و حداقل داده پایلوت را تعریف می‌کند.

۸.۱ اصول مدل داده

- **کانونیکال:** یک زبان مشترک برای همه بیمه‌گران، مستقل از نام فیلدهای داخلی.
- **حداقلی:** فقط فیلدهایی که یک الگو یا یک تصمیم را پشتیبانی می‌کنند.
- **زمانی:** هر رکورد، زمان اعتبار و زمان مشاهده دارد؛ ویژگی‌ها نقطه‌زمانی محاسبه می‌شوند.
- **توکن‌شده:** شناسه‌های مستقیم با توکن جایگزین می‌شوند؛ متن خام در هاب ذخیره نمی‌شود.
- **نسخه‌دار:** هر تغییر در مدل داده، نسخه و نقشه تبدیل دارد.

۸.۲ موجودیت‌های کانونیکال

مدل داده از هفت موجودیت اصلی فراخوان و پنج موجودیت رابطه‌ای تشکیل شده است. نمودار زیر، زنجیره اصلی و گره‌های رابطه‌ای را نشان می‌دهد.



شکل ۸-۱ — مدل گراف زنجیره پرونده و گره‌های رابطه‌ای

جدول ۸-۱ — موجودیت‌های کانونیکال

موجودیت	فیلدهای کلیدی	منبع	کاربرد اصلی
بیمه‌گذار	توکن شخص، بازه سنی، جنسیت، استان، شغل (طبقه)	بیمه‌نامه، CRM	پروفایل ریسک، ارتباط
خودرو	توکن VIN، توکن پلاک، سازنده، مدل، سال، کاربری	بیمه‌نامه	تطبیق خودرو، سابقه خسارت
بیمه‌نامه	توکن بیمه‌نامه، محصول، شروع/پایان، سقف، فرانشیز، پوشش‌ها	سیستم بیمه‌نامه	پوشش، زمان‌بندی، ریسک
حادثه	تاریخ، بازه ساعت، استان/شهر، نوع، شدت، پلیس، مصدومیت	فرم خسارت	تطبیق رخداد، الگوی زمان/مکان
خسارت	توکن خسارت، نوع، علت، مبلغ ادعا/پرداخت (بازه)، وضعیت، چرخه	سیستم خسارت	الگوهای مبلغ، تکرار، شکاف
کارشناس	توکن کارشناس، شرکت، تاریخ ارزیابی، برآورد، تگ‌ها	سیستم ارزیابی	الگوی کارشناس، متن گزارش
پرداخت	توکن پرداخت، مبلغ (بازه)، تاریخ، ذی‌نفع، حساب (توکن)	سیستم پرداخت	الگوی پرداخت، ذی‌نفع تکراری
طرف مقابل	توکن شخص/خودرو، نقش، مصدومیت	فرم خسارت	حلقه تبانی
کارگاه	توکن کارگاه، استان، تخصص	فاکتور/ارزیابی	تمرکز تعمیرگاه
سند	نوع، تاریخ، ویژگی‌های متنی	DMS	NLP، تناقض
رابطه	نوع (تلفن/آدرس/حساب/خانواده)، توکن مرتبط، اطمینان	منابع مجاز	گراف ارتباط
برچسب	نوع تقلب، نتیجه بررسی، تاریخ، شاهد	واحد تقلب	یادگیری و ارزیابی

۸.۳ زنجیره هفت‌گانه فراخوان و پرسشی که هر گره پاسخ می‌دهد

جدول ۸-۲ — زنجیره فراخوان و الگوها

گره	پرسش کلیدی	الگوی فعال‌شده
بیمه‌گذار	چه کسی، با چه سابقه‌ای، در چند شرکت؟	تکرار خسارت، ارتباط افراد
خودرو	همان خودرو، با چه پلاک/VIN، در چند پرونده؟	خسارت تکراری، خودروی پرتکرار
بیمه‌نامه	چه پوششی، با چه زمان‌بندی، نزدیک به حادثه؟	بیمه‌نامه‌های تازه، سوءاستفاده از پوشش
حادثه	چه زمان، مکان، نوع و شدتی؟	تطبیق رخداد، الگوی زمان/مکان
خسارت	چه مبلغی، با چه وضعیتی، در چه چرخه‌ای؟	شکاف ادعا/پرداخت، تکرار مبلغ
کارشناس	چه کسی، با چه برآورد و چه تگ‌هایی؟	الگوی کارشناس، متن گزارش
پرداخت	به چه کسی، چه مبلغی، چه زمانی؟	ذی‌نفع تکراری، حساب مشترک

۸.۴ توکن‌ها و دامنه‌های کلید

هر نوع شناسه، توکن اختصاصی و کلید اختصاصی دارد. جداسازی دامنه‌ها مانع می‌شود که یک توکن در دو کاربرد متفاوت قابل استفاده باشد. توکن‌ها با **SHA256-HMAC** و کلید نگهداری‌شده در HSM تولید می‌شوند؛ خروجی به ۱۲۸ بیت کوتاه می‌شود و با Base32 نمایش داده می‌شود.

جدول ۸-۳ — توکن‌ها و دامنه‌های کلید

توکن	ورودی	دامنه کلید	کاربرد	عمر
subject_T	کدملی نرمال‌شده	subject_K	اتصال پرونده‌های یک بیمه‌گذار	پایدار با نسخه کلید
vehicle_T	VIN یا پلاک نرمال‌شده	vehicle_K	اتصال خودرو	پایدار
policy_T	شناسه بیمه‌نامه + بیمه‌گر	policy_K	یکتایی بیمه‌نامه	دوره بیمه‌نامه
claim_T	شناسه خسارت + بیمه‌گر	claim_K	یکتایی پرونده	دوره نگهداری
event_T	اثر انگشت رخداد	event_K	تطبیق دقیق رخداد	دوره نگهداری
phone_T	شماره تلفن نرمال‌شده	relation_K	گراف ارتباط	دوره نگهداری
address_T	آدرس نرمال‌شده	relation_K	گراف ارتباط	دوره نگهداری
account_T	شناسه حساب/شبا	relation_K	ذی‌نفع مشترک	دوره نگهداری

۸.۵ تطبیق رخداد: از اثر انگشت دقیق تا تطبیق فازی

یک تصادف ممکن است با تاریخ، مبلغ یا شرح متفاوت در دو شرکت ثبت شود. بنابراین یک «توکن رخداد» دقیق کافی نیست. دو مرحله طراحی می‌کنیم: **اثر انگشت دقیق** برای تطبیق قطعی، و **تطبیق فازی** برای پرونده‌های نزدیک.

۱) **بلوک‌بندی:** پرونده‌ها بر اساس استان، بازه تاریخ ±۳ روز و توکن خودرو گروه‌بندی می‌شوند تا فضای مقایسه محدود شود.

۲) **ویژگی‌های تطبیق:** توکن خودرو، توکن طرف مقابل، فاصله تاریخ، فاصله جغرافیایی، شباهت نوع خسارت، نسبت مبلغ، و وجود گزارش پلیس.

۳) **امتیاز شباهت:** ترکیب وزن‌دار ویژگی‌ها؛ امتیاز بالای ۰٫۷۵ «تطبیق بالا»، ۰٫۵۵ تا ۰٫۷۵ «بررسی» و پایین‌تر «غیرمرتبط».

۴) **توضیح:** نمایش ویژگی‌های منطبق و وزن آن‌ها، بدون افشای مبلغ یا تاریخ دقیق شرکت دیگر.

۵) **بازخورد:** تأیید یا رد کارشناس، وزن‌ها را در نسخه بعدی کالیبره می‌کند.

جدول ۸-۴ — ویژگی‌های تطبیق رخداد

ویژگی	وزن پیشنهادی	دلیل
تطبیق توکن خودرو (VIN/پلاک)	۰٫۳۵	قوی‌ترین نشانه وحدت رخداد
تطبیق طرف مقابل / تلفن	۰٫۱۵	نشانه ارتباط مستقیم
فاصله تاریخ (۰ تا ۳ روز)	۰٫۱۵	حوادث نزدیک، احتمال تکرار
فاصله جغرافیایی (استان/شهر)	۰٫۱۰	هم‌مکانی
شباهت نوع و شدت خسارت	۰٫۱۰	الگوی آسیب
نسبت مبلغ ادعا	۰٫۱۰	هم‌ترازی مالی
وجود گزارش پلیس / کروکی	۰٫۰۵	اعتبار رخداد

۸.۶ استانداردسازی داده

جدول ۸-۵ — قواعد استانداردسازی

فیلد	قاعده نرمال‌سازی	نمونه
تاریخ	تبدیل شمسی/میلادی به ISO؛ بازه ساعت ۶ ساعته	03-08-2026 → ۱۴۰۵/۰۵/۱۲
مبلغ	ریال؛ بازه‌بندی لگاریتمی برای تحلیل بین‌شرکتی	۴۸۰،۰۰۰،۰۰۰ → بازه ۴۰۰–۵۰۰ میلیون
جغرافیا	کد استان/شهر استاندارد؛ آدرس فقط توکن	تهران/تهران
خودرو	VIN ۱۷ رقمی؛ پلاک با حذف فاصله و حروف	۱۲B345 → ۳۴۵ب۱۲
نوع خسارت	کد استاندارد داخلی + نقشه به کد کانونیکال	BODY → بدنه
وضعیت	مجموعه ثابت: باز، بسته، رد، پرداخت	—
علت	کد علت + متن آزاد فقط برای NLP	COLLISION → برخورد

۸.۷ کیفیت داده و پذیرش

جدول ۸-۶ — ابعاد کیفیت داده

بُعد	قاعده	آستانه پایلوت	اقدام در نقض
کامل‌بودن	نسبت فیلدهای الزامی موجود	≥ ۹۵٪	قرنطینه رکورد + گزارش به بیمه‌گر
اعتبار	دامنه، فرمت، کد معتبر	≥ ۹۸٪	رد رکورد + بازخورد
یکتایی	عدم تکرار کلید پرونده	۱۰۰٪	حذف تکراری + لاگ
به‌هنگام‌بودن	تأخیر batch	≤ ۲۴ ساعت	هشدار عملیاتی
سازگاری	تطبیق کلیدهای خارجی	≥ ۹۹٪	گزارش شکست اتصال

۸.۸ حداقل داده مورد نیاز پایلوت

جدول زیر، حداقل مجموعه داده برای پایلوت ۱۲ هفته‌ای است. ستون «الزام» مشخص می‌کند که نبود فیلد، مانع شروع است یا فقط کیفیت را کاهش می‌دهد.

جدول ۸-۷ — حداقل داده پایلوت

گروه	فیلد	منبع	الزام	دانه‌بندی	کاربرد
بیمه‌نامه	شناسه بیمه‌نامه (توکن)	بیمه‌نامه	الزامی	هر بیمه‌نامه	اتصال
بیمه‌نامه	محصول و پوشش‌ها	بیمه‌نامه	الزامی	کد	پوشش
بیمه‌نامه	تاریخ شروع/پایان	بیمه‌نامه	الزامی	روز	زمان‌بندی
بیمه‌نامه	سقف و فرانشیز	بیمه‌نامه	مطلوب	بازه	ریسک
بیمه‌گذار	کدملی (برای توکن)	بیمه‌نامه	الزامی	—	توکن موضوعی
بیمه‌گذار	بازه سنی/جنسیت/استان	بیمه‌نامه	مطلوب	بازه	پروفایل
خودرو	VIN یا پلاک	بیمه‌نامه	الزامی	—	تطبیق خودرو
خودرو	سازنده/مدل/سال	بیمه‌نامه	مطلوب	کد	مقایسه
حادثه	تاریخ و ساعت	فرم خسارت	الزامی	روز/بازه	تطبیق رخداد
حادثه	استان/شهر	فرم خسارت	الزامی	کد	جغرافیا
حادثه	نوع و شدت	فرم خسارت	الزامی	کد	الگو
حادثه	گزارش پلیس/کروکی	فرم خسارت	مطلوب	بله/خیر	اعتبار
حادثه	شرح حادثه	فرم خسارت	مطلوب	متن	NLP
خسارت	شناسه خسارت (توکن)	خسارت	الزامی	—	اتصال
خسارت	نوع و علت	خسارت	الزامی	کد	الگو
خسارت	مبلغ ادعا	خسارت	الزامی	بازه	شکاف
خسارت	مبلغ پرداخت/برآورد	خسارت	الزامی	بازه	شکاف
خسارت	وضعیت و تاریخ‌ها	خسارت	الزامی	روز	چرخه
خسارت	علت رد/کاهش	خسارت	مطلوب	کد	برچسب
کارشناسی	شناسه کارشناس/شرکت	ارزیابی	مطلوب	توکن	الگوی کارشناس
کارشناسی	برآورد و تگ‌ها	ارزیابی	الزامی	بازه/کد	الگو
کارشناسی	متن گزارش	ارزیابی	مطلوب	متن	NLP
پرداخت	مبلغ/تاریخ/ذی‌نفع	پرداخت	مطلوب	بازه/روز	الگو
رابطه	تلفن/آدرس/حساب (توکن)	پرداخت/CRM	اختیاری	توکن	گراف
برچسب	نتیجه بررسی تقلب	واحد تقلب	الزامی برای ارزیابی	پرونده	یادگیری

حجم پیشنهادی: حداقل ۲۴ ماه داده تاریخی، ۱۰٬۰۰۰ تا ۳۰٬۰۰۰ پرونده خسارت خودرو، و ترجیحاً ۳۶ ماه برای پوشش فصلی. برای آموزش NLP، حداقل ۵۰۰ گزارش کارشناسی با برچسب کیفی لازم است. اگر برچسب تقلب کافی نباشد، از ترکیب پرونده‌های تأییدشده، برچسب ضعیف و شبیه‌سازی استفاده می‌شود؛ اما نتایج باید با احتیاط گزارش شوند.

۸.۹ تبار داده و فراداده

هر رکورد، مسیر تبدیل خود را از سیستم منبع تا ویژگی نهایی حفظ می‌کند: نام فیلد مبدأ، قاعده تبدیل، نسخه مدل داده، زمان پردازش و اپراتور. این تبار داده برای سه کار لازم است: پاسخ به ممیزی، بازتولید نتیجه یک پرونده در گذشته، و تشخیص خطای یکپارچه‌سازی. مخزن فراداده، مستقل از داده تحلیلی و با دسترسی خواندنی برای ناظر نگهداری می‌شود.

۹ موتور بی‌نام‌سازی و حریم خصوصی

بی‌نام‌سازی یک تابع نیست؛ یک زنجیره کنترل است. این فصل، مدل تهدید، طراحی توکن، نرمال‌سازی، حداقل‌سازی، کنترل استعلام، DPIA و مرزهای ادعای ما را مشخص می‌کند. جمله کلیدی: ما ادعای «ناشناس مطلق» نمی‌کنیم؛ ادعا می‌کنیم دسترسی به داده خام و بازشناسی، از نظر طراحی و کنترل، غیرممکن یا پرهزینه می‌شود.

۹.۱ مدل تهدید

جدول ۹-۱ — مدل تهدید و کنترل‌ها

تهدید	سناریو	اثر	کنترل
نشت توسط کاربر داخلی	کارشناس، داده یک فرد مشهور را جست‌وجو می‌کند	بالا	RBAC، ثبت استعلام، حداقل افشا، هشدار
سوءاستفاده از استعلام	بیمه‌گر با استعلام انبوه، الگوی بازار را بازسازی می‌کند	بالا	سقف نرخ، سهمیه، تشخیص الگو
حمله پیوند	ترکیب داده با منبع بیرونی برای بازشناسی	بالا	حداقل‌سازی، بازه‌بندی، k-آنونیمیتی
وارونگی مدل	استنتاج داده آموزش از خروجی مدل	متوسط	عدم خروجی خام، محدودیت API
نشت زیرساخت	دسترسی به پایگاه داده یا پشتیبان	بالا	رمزنگاری، HSM، تفکیک شبکه
استنتاج از تجمیع	گزارش‌های کوچک، یک شرکت را افشا می‌کند	متوسط	حداقل اندازه گروه، نویز، سرکوب

۹.۲ طراحی توکن‌سازی

کدملی ایران ۱۰ رقم است و فضای آن قابل پیمایش؛ بنابراین هش ساده یا نمک عمومی کافی نیست. مهاجم با یک فرهنگ لغت از کدهای ملی می‌تواند هش‌ها را بازسازی کند. راه‌حل، **HMAC با کلید مخفی** است که در HSM نگهداری می‌شود و هرگز به تحلیل‌گر یا بیمه‌گر داده نمی‌شود.

جدول ۹-۲ — کنترل‌های توکن‌سازی

کنترل	پیاده‌سازی	نحوه تأیید
تابع کلیددار	SHA256-HMAC با کلید ۲۵۶ بیتی	بازبینی کد و آزمون بردار
کوتاه‌سازی	۱۲۸ بیت، Base32	آزمون یکتایی و تصادف
دامنه کلید	کلید جدا برای شخص، خودرو، رخداد، رابطه	بررسی پیکربندی
نگهداری کلید	HSM یا KMS با تفکیک وظایف	ممیزی دسترسی
چرخش کلید	نسخه‌بندی توکن، حفظ کلید قدیم برای اتصال	آزمون مهاجرت
عدم ذخیره	شناسه خام در لاگ و جدول ذخیره نمی‌شود	اسکن لاگ و آزمون نفوذ
نرخ محدود	سقف توکن‌سازی و استعلام	پایش و هشدار

۹.۳ نرمال‌سازی قبل از توکن

اگر یک بیمه‌گر «۰۱۲۳۴۵۶۷۸۹» و دیگری «۱۲۳۴۵۶۷۸۹» بفرستد، توکن‌ها متفاوت می‌شوند و تطبیق شکست می‌خورد. نرمال‌سازی، پیش‌نیاز تطبیق است و باید در سرویس مرکزی و نه در بیمه‌گر انجام شود تا یکسان باشد.

جدول ۹-۳ — نرمال‌سازی شناسه‌ها

ورودی	نرمال‌سازی	ریسک
کدملی	حذف فاصله، تبدیل ارقام فارسی/عربی به لاتین، اعتبارسنجی چک‌سام	کد نامعتبر، صفر ابتدایی
پلاک	حذف فاصله و خط تیره، یکسان‌سازی حروف	پلاک تعویض‌شده
VIN	حذف نویسه‌های غیرمجاز، بزرگ‌کردن حروف	VIN ناقص
تلفن	حذف پیشوند، یکسان‌سازی ۰۹	شماره مشترک
آدرس	یکسان‌سازی واژگان، حذف Stopword، توکن	آدرس ناقص
حساب/شبا	حذف فاصله، اعتبارسنجی IBAN	حساب مشترک

۹.۴ حداقل‌سازی و تعمیم

جدول ۹-۴ — حداقل‌سازی داده

فیلد	داده خام (نزد بیمه‌گر)	داده مشترک (هاب)	دلیل
سن	تاریخ تولد کامل	بازه ۵ ساله	کاهش بازشناسی
مبلغ	مبلغ دقیق	بازه لگاریتمی	کاهش افشای مالی
تاریخ	تاریخ و ساعت دقیق	روز + بازه ۶ ساعته	تطبیق رخداد با حداقل دقت
جغرافیا	آدرس کامل	استان/شهر	الگوی مکانی
نام و کدملی	کامل	توکن	حذف هویت
متن گزارش	متن کامل	ویژگی و برچسب	بدون ذخیره متن NLP
حساب بانکی	شبا کامل	توکن	کشف ذی‌نفع مشترک

۹.۵ کنترل‌های استعلام و پاسخ

جدول ۹-۵ — کنترل‌های استعلام

کنترل	قاعده	اجرا
هدف پردازش	هر استعلام باید کد هدف معتبر داشته باشد	Policy engine
حداقل افشا	پاسخ فقط امتیاز، برچسب، کد دلیل	قرارداد API
ممنوعیت نام شرکت	هیچ پاسخ یا لاگی نام رقیب را برنمی‌گرداند	طراحی و آزمون
ممنوعیت مبلغ دقیق	فقط بازه یا نسبت	تبدیل در لایه سرویس
سقف نرخ	حداکثر استعلام در دقیقه/روز	API Gateway
حداقل گروه	گزارش تجمیعی با $k \geq 5$	لایه گزارش
پایش الگو	استعلام انبوه یا مکرر، هشدار	SIEM
ثبت کامل	چه کسی، چه زمانی، چه هدفی	Audit log

۹.۶ حریم خصوصی تفاضلی و مدل فدرال (نقشه راه)

برای گزارش‌های تجمیعی و بنچمارک، می‌توان از **حریم خصوصی تفاضلی** با نویز کنترل‌شده و بودجه مشخص استفاده کرد. برای آموزش مدل بدون انتقال داده، **یادگیری فدرال** و برای تطبیق بدون اشتراک توکن، **اشتراک خصوصی مجموعه‌ها (PSI)** گزینه‌های آینده هستند. این فناوری‌ها در MVP لازم نیستند و عامدانه وارد دامنه نشده‌اند، زیرا پیچیدگی اجرایی و زمان پایلوت را افزایش می‌دهند؛ اما معماری برای افزودن آن‌ها باز است.

> **چرا این فناوری‌ها در MVP نیستند**
>
> پایلوت ۱۲ هفته‌ای باید یک نتیجه قابل دفاع بدهد، نه یک نمایش فناوری. توکن‌سازی کلیددار + حداقل‌سازی + کنترل استعلام، برای اثبات مفهوم کافی است. افزودن رمزنگاری همریخت یا یادگیری فدرال در فاز اول، ریسک تأخیر را بدون افزایش متناسب دقت ایجاد می‌کند.

۹.۷ ارزیابی اثر حریم خصوصی (DPIA)

جدول ۹-۶ — گام‌های DPIA

گام	فعالیت	خروجی
۱. توصیف پردازش	هدف، داده، ذی‌نفعان، جریان داده	نقشه پردازش
۲. ضرورت و تناسب	آیا داده کمتری کافی است؟	توجیه حداقل‌سازی
۳. شناسایی ریسک	بازشناسی، نشت، سوءاستفاده	ثبت ریسک
۴. کنترل	توکن، دسترسی، نرخ، ممیزی	برنامه کنترل
۵. تأیید	DPO، حقوقی، ناظر	امضا و تاریخ بازبینی

۹.۸ مبنای قانونی، رضایت و حقوق افراد

پیش از پایلوت باید سه پرسش حقوقی پاسخ داده شود: **مبنای پردازش** چیست (الزام قانونی، اجرای قرارداد، منافع مشروع یا رضایت)؟ **نقش‌ها** چگونه تقسیم می‌شوند (مسئول پردازش، پردازشگر، مشترک)؟ و **مدت نگهداری** چقدر است؟ پاسخ‌ها در یک «ثبت هدف و مبنا» در سامانه ذخیره می‌شود و هر استعلام به یک هدف

معتبر متصل است. حقوق افراد — اطلاع، دسترسی، اصلاح، اعتراض — از طریق بیمه‌گر و با پاسخ‌گویی مستند اجرا می‌شود.

از منظر انطباق، این طرح به قوانین و مقررات حاکم از جمله قانون بیمه و آیین‌نامه‌های بیمه مرکزی، قانون تجارت الکترونیک، قانون جرایم رایانه‌ای، الزامات افتا و مصوبات شورای عالی فضای مجازی متکی است. در صورت ابلاغ قانون حفاظت از داده‌های شخصی، الزامات آن نیز در DPIA و قراردادها اعمال می‌شود. جزئیات در فصل ۱۴ آمده است.

۹.۹ مرز ادعا: شبه‌بی‌نام‌سازی، ناشناس‌سازی نیست

توکن پایدار، رکوردهای یک فرد را در چند شرکت به هم متصل می‌کند. بنابراین داده شبه‌بی‌نام، همچنان **داده شخصی** است و مشمول الزامات حفاظت می‌شود. ما این واقعیت را پنهان نمی‌کنیم. ادعای ما سه چیز است: نخست، داده خام هویتی در هاب ذخیره نمی‌شود؛ دوم، دسترسی به نگاشت هویت و کلید، از نظر فنی و سازمانی جدا است؛ سوم، هر استعلام ثبت و محدود می‌شود. این سه، ریسک را به سطح قابل مدیریت کاهش می‌دهد، اما جای ارزیابی حقوقی و DPIA را نمی‌گیرد.

> **جمله‌ای که در قرارداد می‌آید**
>
> «داده‌های تبادل‌شده شبه‌بی‌نام و حداقلی هستند؛ طرفین متعهد به عدم بازشناسی، عدم ترکیب با منابع بیرونی و عدم استفاده برای هدفی جز کشف تقلب می‌شوند.»

۱۰ یکپارچه‌سازی و اتصال به سیستم‌های بیمه

بیشترین علت شکست پروژه‌های داده در بیمه، نه الگوریتم است و نه زیرساخت؛ «اصطکاک اتصال» است. این فصل، الگوهای اتصال، نقشه سیستم‌ها، طراحی API، فرایند onboarding و سناریوی پایلوت را مشخص می‌کند. اصل حاکم: حداقل تغییر در سیستم بیمه‌گر، حداکثر ارزش در لایه تحلیل.

۱۰.۱ اصل: batch اول، API بعد

در پایلوت، هیچ بیمه‌گری نباید مجبور شود سیستم خسارت خود را تغییر دهد. مسیر پیشنهادی، **بارگذاری دسته‌ای شبانه** برای داده تاریخی و روزانه، و **API استعلام** برای پرونده‌های جاری است. جریان رخداد (Kafka) و CDC در فازهای بعدی و فقط برای بیمه‌گران با بلوغ فنی بالا پیشنهاد می‌شود. این ترتیب، زمان اتصال را از ماه‌ها به هفته‌ها کاهش می‌دهد.

۱۰.۲ الگوهای اتصال

جدول ۱۰-۱ — الگوهای اتصال

الگو	کاربرد	تأخیر	پیچیدگی	پیشنهاد
بارگذاری دسته‌ای (PGP + SFTP)	داده تاریخی و روزانه	ساعت	کم	الزامی برای پایلوت
API استعلام (REST/JSON)	بررسی پرونده جاری	ثانیه	متوسط	الزامی برای پایلوت
جریان رخداد (Kafka)	هشدار لحظه‌ای	میلی‌ثانیه	بالا	اختیاری فاز ۲
CDC پایگاه داده	همگام‌سازی خودکار	دقیقه	بالا	اختیاری فاز ۲
فایل دستی / اکسل	شروع سریع	روز	بسیار کم	فقط نمونه اولیه

۱۰.۳ نقشه سیستم‌های بیمه‌گر

جدول ۱۰-۲ — سیستم‌های منبع و مسئولیت

سیستم	داده	جهت	روش	مالک
بیمه‌نامه	پوشش، سقف، فرانشیز، insured	خروجی	API / batch	مدیریت بیمه‌نامه
خسارت	پرونده، مبلغ، وضعیت	خروجی + استعلام	API / batch	مدیریت خسارت
ارزیابی	برآورد، تگ، گزارش	خروجی	API / batch	مدیریت ارزیابی
پرداخت	مبلغ، تاریخ، ذی‌نفع	خروجی	batch	مالی
CRM	مشتری، تماس، آدرس	خروجی	batch	بازاریابی
DMS	اسناد و متن	خروجی	API / batch	بایگانی
واحد تقلب	برچسب و نتیجه بررسی	ورودی + خروجی	فایل / API	مدیریت تقلب

۱۰.۴ طراحی API

API بـا **JSON/REST**، احـراز هـویت **mTLS + OAuth2 Client Credentials**، شـناسـه درخـواسـت یکتا (Key-Idempotency)، نسخه‌بندی مسیر (`/v1/...`) و محدودیت نرخ طراحی می‌شود. هیچ پاسخ API شامل شناسه خام، نام شرکت دیگر، مبلغ دقیق یا تاریخ رخداد رقیب نیست.

جدول ۱۰-۳ —مسیرهای اصلی API

مسیر	متد	کاربرد	احراز هویت	SLA
v1/tokenize/	POST	تبدیل شناسه به توکن (فقط سرویس مجاز)	OAuth2 + mTLS	۳۰۰ms > p95
v1/claims/	POST	ثبت/به‌روزرسانی رخداد خسارت	OAuth2	۵۰۰ms > p95
v1/score/	POST	دریافت امتیاز و کد دلیل برای یک پرونده	OAuth2	۲s > p95
v1/cases/	GET	صف اولویت و جزئیات پرونده	OAuth2	۱s > p95
v1/feedback/	POST	ثبت نتیجه بررسی کارشناس	OAuth2	۵۰۰ms > p95
v1/health/	GET	پایش سلامت سرویس	mTLS	—

نمونه کامل درخواست و پاسخ، کدهای خطا و قرارداد داده در **پیوست ب** آمده است. پاسخ نمونه شامل `score_risk`، `score_health`، `badges`، `codes_reason`، `confidence` و `version_rules` است؛ هیچ فیلد هویتی یا مالی دقیق در پاسخ نیست.

۱۰.۵ فرایند onboarding بیمه‌گر

جدول ۱۰-۴ — فرایند onboarding

گام	خروجی	مسئول	مدت
۱. جلسه شروع	دامنه، نقش‌ها، تقویم	PMO	۱ روز
۲. نقشه‌برداری داده	نقشه فیلد مبدأ به کانونیکال	داده + بیمه‌گر	۳ روز
۳. نمونه داده	فایل نمونه بی‌نام/شبه‌واقعی	بیمه‌گر	۳ روز
۴. کیفیت داده	گزارش DQ و شکاف‌ها	تیم داده	۲ روز
۵. بازبینی امنیتی	چک‌لیست امنیت و انطباق	امنیت	۲ روز
۶. محیط sandbox	دسترسی و کلید تست	عملیات	۲ روز
۷. اتصال	فعال API/batch	تیم فنی	۵ روز
۸. آزمون یکپارچه	نتایج آزمون و رفع خطا	QA	۳ روز
۹. UAT	تأیید کارشناس	بیمه‌گر	۳ روز
۱۰. hypercare و live-go	راه‌اندازی و پشتیبانی نزدیک	همه	۵ روز

۱۰.۶ کنترل کیفیت در نقطه ورود

- **اعتبارسنجی طرح:** نوع فیلد، طول، دامنه و کدهای مجاز پیش از پذیرش.
- **یکپارچگی ارجاعی:** وجود بیمه‌نامه/خسارت/خودرو مرتبط.

- **تشخیص تکرار:** کلید پرونده و ترکیب بیمه‌نامه + تاریخ.
- **اسکن PII:** شناسه خام در فیلدهای غیرمجاز، لاگ یا نام فایل.
- **قرنطینه:** رکورد نامعتبر رد و با کد خطا به بیمه‌گر برگردانده می‌شود.
- **گزارش:** امتیاز کیفیت هر بیمه‌گر در داشبورد عملیاتی.

۱۰.۷ کارایی و مقیاس

برای پایلوت، یک محیط با ۸ هسته CPU، ۳۲ گیگابایت RAM و ۱ ترابایت SSD کافی است. برای تولید ۱۰ میلیون پرونده و ۵۰ بیمه‌گر، معماری افقی با پایگاه داده تحلیل ستونی (ClickHouse) و ذخیره شیء (S3/MinIO) پیشنهاد می‌شود. ظرفیت‌سنجی نهایی پس از تحلیل حجم داده دو بیمه‌گر پایلوت انجام می‌شود.

۱۰.۸ انطباق با استانداردهای داده صنعت

مدل کانونیکال به‌گونه‌ای طراحی می‌شود که به استانداردهای داده بیمه مرکزی و کدهای رایج صنعت نگاشت شود. در صورت وجود استاندارد اجباری برای تبادل داده، لایه نگاشت جداگانه اضافه می‌شود و مدل داخلی تغییر نمی‌کند. هدف، جلوگیری از قفل‌شدن به یک نسخه خاص و تسهیل اتصال بیمه‌گران جدید است.

۱۰.۹ سناریوی اتصال پایلوت

جدول ۱۰-۵ — سناریوی اتصال پایلوت

هفته	بیمه‌گر A	بیمه‌گر B	آزمایشگاه داده
۱ تا ۲	توافق، نقشه داده، نمونه	توافق، نقشه داده، نمونه	محیط و دسترسی
۳ تا ۴	batch + DQ تاریخی	batch + DQ تاریخی	بارگذاری داده مرجع
۵ تا ۶	API استعلام در sandbox	batch روزانه	پایش کیفیت
۷ تا ۹	اتصال میز کار	اتصال گزارش	اعتبارسنجی گذشته‌نگر
۱۰ تا ۱۲	پایلوت سایه	پایلوت سایه	گزارش نهایی

۱۱ موتور کشف تقلب: الگوها، گراف و متن

موتور تشخیص، قلب تحلیلی راهکار است. طراحی آن عامدانه ترکیبی است: قواعد صریح برای شفافیت، گراف برای رابطه، ناهنجاری برای کشف الگوهای ناشناخته، و NLP برای متن گزارش. هیچ‌کدام به‌تنهایی تصمیم نمی‌گیرد؛ هر کدام یک «شاهد» تولید می‌کند.

۱۱.۱ معماری پنج‌لایه تشخیص

جدول ۱۱-۱ — لایه‌های تشخیص

لایه	ورودی	خروجی	نقش
۱. قواعد صریح	فیلدهای کانونیکال	نشانه قطعی/احتمالی	شفافیت و سرعت
۲. گراف ارتباط	گره‌ها و یال‌های توکن‌شده	خوشه، مسیر، حلقه	کشف رابطه پنهان
۳. ناهنجاری	ویژگی‌های عددی	امتیاز انحراف + ویژگی مؤثر	کشف الگوی ناشناخته
۴. NLP	متن گزارش/شرح	برچسب و شاهد متنی	کشف اغراق/تناقض
۵. ادغام و امتیاز	شواهد لایه‌ها	امتیاز، کد دلیل، اطمینان	اولویت‌بندی

۱۱.۲ الگوهای اصلی و مکمل

فراخوان ۵ هشت خانواده الگو را نام می‌برد. ما آن‌ها را به **چهار الگوی اصلی** (بیشترین اثر و داده) و **پنج الگوی مکمل** (تقویت‌کننده) تقسیم می‌کنیم. این تقسیم، دامنه MVP را کنترل می‌کند و در عین حال پوشش کامل فراخوان را حفظ می‌کند.

الگوی اصلی ۱ — خسارت تکراری و چندشرکتی (Insurer-Multi / Duplicate)

این الگو تلاش برای تشکیل پرونده و دریافت خسارت یک تصادف از بیش از یک شرکت را شناسایی می‌کند. موتور، توکن بیمه‌گذار و توکن خودرو را با **تطبیق رخداد** (اثر انگشت دقیق + شباهت فازی) ترکیب می‌کند. وجود بیش از یک پرونده فعال برای رخدادی واحد، نشانه‌ای برای بررسی است؛ بدون افشای هویت یا نام شرکت‌های درگیر. امتیاز پیشنهادی: تا ۲۵ نمره برای تطبیق بالا؛ ۱۵ نمره برای تطبیق متوسط.

الگوی اصلی ۲ — شکاف ادعا تا واقعیت کارشناسی (Claim vs. Settlement)

برای نمونه، بیمه‌گذار خسارت ۵۰۰ میلیونی اعلام می‌کند و ارزیابی کارشناسی آن را به ۳۰۰ میلیون کاهش می‌دهد. یک اختلاف منفرد به‌تنهایی نشانه تقلب نیست؛ تکرار اختلاف معنادار میان مبلغ ادعا و مبلغ نهایی در چند پرونده، می‌تواند نشانه بزرگ‌نمایی سیستماتیک باشد. نسبت `(ادعا − پرداخت) / ادعا` در بازه‌های زمانی و نوع خسارت محاسبه می‌شود. امتیاز: تا ۱۲ نمره.

الگوی اصلی ۳ — سوابق صحنه‌سازی و خسارت عمدی

تگ‌ها و هشدارهایی که کارشناسان در پرونده‌های قبلی درباره صحنه‌سازی، صوری‌بودن تصادف یا ایجاد عمدی خسارت ثبت کرده‌اند، پس از اعتبارسنجی به نشانه‌ای برای بررسی تبدیل می‌شوند. تکرار تگ‌های مستقل در بازه زمانی تعریف‌شده، بر امتیاز اثر می‌گذارد؛ یک یادداشت منفرد حکم قطعی نیست. امتیاز: تا ۲۰ نمره برای تگ تأییدشده، ۱۰ نمره برای تگ مشکوک.

الگوی اصلی ۴ — تحلیل متنی گزارش کارشناسی (NLP)

پردازش زبان طبیعی روی متن گزارش‌ها می‌تواند واژگان و عبارت‌های حاکی از اغراق، ابهام در تعیین مقصر، تناقض زمانی/مکانی یا شائبه تبانی را بیابد. تحلیل متن صرفاً نشانه و برچسب تولید می‌کند؛ اثر آن بر امتیاز باید پس از عبور از آستانه اطمینان و بازبینی انسانی اعمال شود. امتیاز: تا ۱۰ نمره، مشروط.

الگوی مکمل ۵ — اطلاعات خلاف واقع و ناسازگاری هویت

ناسازگاری در پلاک/VIN، تاریخ‌ها، نام مالک، شماره تماس یا آدرس، در ترکیب با سایر نشانه‌ها بررسی می‌شود. این الگو به‌تنهایی امتیاز بالا نمی‌دهد؛ زیرا خطای ثبت داده نیز می‌تواند همین نشانه را ایجاد کند. امتیاز: تا ۱۵ نمره با تأیید.

الگوی مکمل ۶ — ارتباطات غیرعادی و حلقه تبانی

اشتراک تلفن، آدرس، حساب بانکی، کارگاه، کارشناس یا طرف مقابل میان چند پرونده، می‌تواند نشانه یک شبکه باشد. گراف، خوشه‌های متراکم و مسیرهای تکرارشونده را پیدا می‌کند. برای جلوگیری از «اتهام به‌واسطه ارتباط»، تنها خوشه‌هایی با چند شاهد مستقل و اندازه کافی بررسی می‌شوند. امتیاز: تا ۲۰ نمره.

الگوی مکمل ۷ — تکرار الگو در زمان، مبلغ یا نوع خسارت

تکرار خسارت با فاصله زمانی کوتاه، مبلغ مشابه یا نوع آسیب یکسان، در ترکیب با سابقه بیمه‌نامه بررسی می‌شود. این الگو به کشف «الگوی ساختگی» کمک می‌کند، اما به‌تنهایی ضعیف است. امتیاز: تا ۱۰ نمره.

الگوی مکمل ۸ — سوابق غیرعادی خسارت یا کارشناسی

تعداد بالای خسارت در بازه کوتاه، نسبت غیرعادی خسارت به حق بیمه، یا تمرکز غیرعادی یک کارشناس/کارگاه روی یک بیمه‌گذار، نشانه‌های آماری هستند. این الگو بیشتر برای اولویت‌بندی و بررسی تجمیعی استفاده می‌شود. امتیاز: تا ۱۰ نمره.

الگوی مکمل ۹ — دستکاری اسناد و مدارک

نشانه‌های متنی یا فراداده‌ای مانند تکرار قالب، تاریخ‌های ناسازگار، تصویر تکراری یا فاکتور با الگوی مشابه، در صورت وجود داده، بررسی می‌شود. در MVP، این الگو فقط در سطح هشدار است و امتیاز مستقیم ندارد؛ زیرا نیازمند تحلیل تصویر و فراداده است.

جدول ۱۱-۲ — خلاصه الگوها و وزن‌ها

الگو	نوع	قاعده نمونه	حداکثر کسر	ریسک مثبت کاذب
خسارت هم‌زمان/چندشرکتی	اصلی	بیش از یک پرونده برای یک رخداد با تطبیق ≤ ۰٫۷۵	۲۵	کم
شکاف ادعا/پرداخت	اصلی	اختلاف ≤ ۳۰٪ در ۲ پرونده از ۳ پرونده اخیر	۱۲	متوسط
صحنه‌سازی تأییدشده	اصلی	تگ تأییدشده در ۲ پرونده مستقل	۲۰	کم
نشانه متنی	اصلی	دو نشانه با اطمینان ≤ ۰٫۸	۱۰	متوسط
اطلاعات خلاف واقع	مکمل	ناسازگاری VIN/پلاک با سابقه	۱۵	متوسط
حلقه ارتباطی	مکمل	خوشه ≤ ۵ پرونده، ≤ ۳ بیمه‌گر، ≤ ۲ شاهد رابطه	۲۰	متوسط
تکرار زمان/مبلغ/نوع	مکمل	۳ خسارت در ۹۰ روز با الگوی مشابه	۱۰	بالا
سابقه غیرعادی	مکمل	نرخ خسارت بالای صدک ۹۵ هم‌گروه	۱۰	بالا
دستکاری سند	مکمل	نشانه فراداده‌ای/تصویری	۰ (هشدار)	بالا

۱۱.۳ تحلیل گراف

گراف، تفاوت اصلی این راهکار با موتورهای قواعد سنتی است. گره‌ها شامل بیمه‌گذار، خودرو، بیمه‌نامه، حادثه، خسارت، کارشناس، پرداخت، کارگاه، طرف مقابل و شناسه‌های رابطه‌ای هستند. یال‌ها، رابطه و زمان اعتبار دارند. تحلیل روی **گراف نقطه‌زمانی** انجام می‌شود؛ یعنی فقط یال‌هایی که پیش از تاریخ اعلام خسارت وجود داشته‌اند.

جدول ۱۱-۳ — الگوریتم‌های گراف

الگوریتم	پرسش	خروجی	کاربرد
مؤلفه‌های متصل	چه پرونده‌هایی به هم وصل‌اند؟	خوشه	کشف شبکه
Louvain	خوشه‌های متراکم کدام‌اند؟	جامعه	حلقه تبانی
مرکزیت بینابینی	کدام گره واسطه است؟	رتبه	کارگاه/کارشناس کلیدی
موتیف زمانی	چه الگوی تکراری در زمان شکل می‌گیرد؟	الگو	تکرار ساختگی
شباهت گره	چه گره‌هایی شبیه‌اند؟	جفت مشابه	رابطه پنهان
انتشار ریسک	ریسک از پرونده تأییدشده چقدر منتشر می‌شود؟	امتیاز زمینه	اولویت‌بندی

قاعده حلقه تبانی: خوشه‌ای با حداقل ۵ پرونده، ۳ بیمه‌گذار و ۲ شاهد رابطه مستقل (مثلاً تلفن مشترک + کارگاه مشترک) و تراکم بالای آستانه، برای بررسی دستی علامت‌گذاری می‌شود. خروجی، **حداقل زیرگراف مسیر** است: «خسارت A ← خودرو V ← تلفن P ← خسارت B»؛ بدون نام، مبلغ یا تاریخ دقیق.

۱۱.۴ کشف ناهنجاری

کشف ناهنجاری برای الگوهایی است که در قواعد صریح نمی‌گنجند. در MVP از روش‌های ساده و قابل توضیح استفاده می‌کنیم: **score-z مقاوم**، **Isolation Forest** و **HBOS**. مدل‌های عمیق در فاز اول عامدانه کنار گذاشته می‌شوند، زیرا داده برچسب‌دار کم و توضیح‌پذیری پایین دارند.

جدول ۱۱-۴ — ویژگی‌های ناهنجاری

ویژگی	نوع ناهنجاری	مثال
نسبت مبلغ ادعا به سقف پوشش	مبلغی	ادعای ۹۵٪ سقف در بیمه‌نامه تازه
فاصله شروع بیمه‌نامه تا حادثه	زمانی	حادثه ۳ روز پس از شروع
تعداد خسارت در ۱۲ ماه	تکرار	۴ خسارت در یک خودرو
نسبت خسارت به حق بیمه	اقتصادی	نسبت ۳ برابر هم‌گروه
تمرکز کارگاه/کارشناس	شبکه‌ای	۶۰٪ پرونده‌ها با یک کارگاه
اختلاف برآورد کارشناس و پرداخت	فرآیندی	اختلاف ۴۰٪

۱۱.۵ تحلیل متن فارسی گزارش کارشناسی

گزارش‌های کارشناسی، منبعی غنی و کم‌استفاده‌اند. اما NLP در این حوزه پرخطر است: یک واژه مبهم یا یک جمله ناقص نباید به اتهام تبدیل شود. طراحی ما، **استخراج شاهد** است، نه صدور حکم. متن خام در هاب ذخیره نمی‌شود؛ سرویس NLP به‌صورت بی‌حالت (stateless) متن را پردازش و فقط ویژگی/برچسب برمی‌گرداند.

جدول ۱۱-۵ — خط لوله NLP

مرحله	روش	خروجی	دروازه کیفیت
OCR	PaddleOCR/Tesseract فارسی	متن از تصویر	دقت ≥ ۹۰٪ روی نمونه
نرمال‌سازی	یکسان‌سازی حروف/ارقام/فاصله	متن تمیز	پوشش ≥ ۹۹٪
قطعه‌بندی	تشخیص بخش‌های گزارش	بخش‌ها	دقت ≥ ۹۰٪
استخراج موجودیت	قواعد + مدل	تاریخ، مبلغ، قطعه، مکان	F1 ≥ ۰٫۸۵
طبقه‌بندی جمله	ParsBERT + IDF-TF + Lexicon	برچسب ریسک	دقت ≥ ۰٫۸ در آستانه بالا
تجمیع	قواعد اطمینان	برچسب گزارش + شاهد	بازبینی انسانی

برچسب‌ها: اغراق در خسارت، ابهام در تعیین مقصر، تناقض زمانی/مکانی، نشانه تبانی، توصیف غیرعادی صحنه، و عدم تطابق با آسیب. **شاهد:** جمله و بازه متن، نه نمره صرف. **آستانه:** فقط برچسب با اطمینان بالا (≤ ۰٫۸) به امتیاز وارد می‌شود؛ بقیه به‌عنوان «اطلاعات کمکی» نمایش داده می‌شوند.

۱۱.۶ کاتالوگ قواعد

هر قاعده یک شناسه، تعریف، شرط، وزن، سطح اطمینان، شاهد و نسخه دارد. قواعد در فایل نسخه‌دار (Git) نگهداری و پیش از انتشار، روی داده تاریخی آزمون می‌شوند. جدول زیر نمونه‌ای از کاتالوگ است؛ کاتالوگ کامل در پیوست ج آمده است.

جدول ۱۱-۶ — نمونه کاتالوگ قواعد

کد	الگو	شرط نمونه	وزن	شاهد
01-DUP	خسارت هم‌زمان	تطبیق رخداد ≤ ۰٫۸۵ بین دو بیمه‌گر	۲۵	ویژگی‌های تطبیق
02-DUP	خسارت تکراری داخلی	دو پرونده با VIN یکسان در ۷ روز	۲۰	VIN + تاریخ
03-DUP	چندپرونده‌ای	۳ پرونده با توکن خودرو یکسان در ۹۰ روز	۱۵	توکن خودرو
01-GAP	شکاف ادعا/پرداخت	اختلاف ≤ ۳۰٪ در ۲ پرونده از ۳	۱۲	نسبت‌ها
02-GAP	بزرگ‌نمایی مکرر	میانگین اختلاف ≤ ۲۵٪ در ۱۲ ماه	۱۰	روند
01-STG	صحنه‌سازی تأییدشده	تگ تأییدشده در ۲ پرونده مستقل	۲۰	تگ کارشناس
02-STG	تگ مشکوک	تگ مشکوک در ۳ پرونده	۱۰	تگ
01-NLP	نشانه متنی قوی	۲ برچسب با اطمینان ≤ ۰٫۸	۱۰	جمله شاهد
02-NLP	تناقض زمانی	تاریخ گزارش با حادثه ناسازگار	۸	متن
01-REL	تلفن مشترک	تلفن یکسان در ≤ ۳ پرونده	۱۵	گراف
02-REL	کارگاه متمرکز	≤ ۵۰٪ پرونده‌ها با یک کارگاه	۱۲	گراف
03-REL	حلقه تبانی	خوشه ≤ ۵ پرونده، ≤ ۲ شاهد	۲۰	زیرگراف
01-HIS	سابقه پرخسارت	≤ ۴ خسارت در ۱۲ ماه	۱۰	شمارش
01-AMT	مبلغ غیرعادی	score-z ≤ ۳ نسبت به هم‌گروه	۸	توزیع
01-ID	ناسازگاری هویت	VIN/پلاک با سابقه ناسازگار	۱۵	تطبیق
01-DOC	سند مشکوک	فراداده/تصویر تکراری	۰ (هشدار)	فراداده

۱۱.۷ ادغام سیگنال‌ها و جلوگیری از شمارش مضاعف

چند قاعده می‌توانند یک واقعیت را توصیف کنند؛ مثلاً «خسارت هم‌زمان» و «تکرار خودرو» ممکن است به یک رخداد اشاره کنند. برای جلوگیری از شمارش مضاعف، سه کنترل اعمال می‌شود: **وابستگی قاعده** (اگر یک شاهد سخت وجود دارد، قواعد مشتق‌شده وزن کمتری می‌گیرند)، **هم‌بستگی** (قواعد با ورودی مشترک، در یک گروه وزن‌دهی می‌شوند)، و **سقف** (جمع کسرها از ۱۰۰ بیشتر نمی‌شود).

۱۱.۸ حلقه بازخورد و یادگیری

نتیجه بررسی کارشناس به‌صورت برچسب کنترل‌شده بازمی‌گردد: تأیید تقلب، رد، نیاز به بررسی بیشتر، یا داده ناکافی. این برچسب‌ها برای کالیبراسیون وزن‌ها، بهبود NLP و اندازه‌گیری دقت استفاده می‌شوند. چرخه به‌روزرسانی **ماهانه** برای وزن‌ها و **فصلی** برای مدل‌های NLP پیشنهاد می‌شود؛ هر تغییر، نسخه جدید و آزمون پس‌رو دارد.

۱۱.۹ محدودیت‌ها: چه چیزی را نمی‌سازیم

- **دروغ‌یاب نیستیم:** مدل، احتمال را اولویت‌بندی می‌کند؛ واقعیت را کارشناس تأیید می‌کند.

- **حکم قطعی صادر نمی‌کنیم:** هیچ پرونده‌ای فقط بر اساس امتیاز رد نمی‌شود.
- **جایگزین بررسی میدانی نیستیم:** گراف و متن، شواهد را کنار هم می‌گذارند، نه جایگزین بازرسی.
- **همه تقلب را کشف نمی‌کنیم:** الگوهای بدون داده یا بدون برچسب، خارج از دامنه MVP هستند.
- **از داده حساس غیرضروری استفاده نمی‌کنیم:** هر فیلد باید هدف و مجوز داشته باشد.

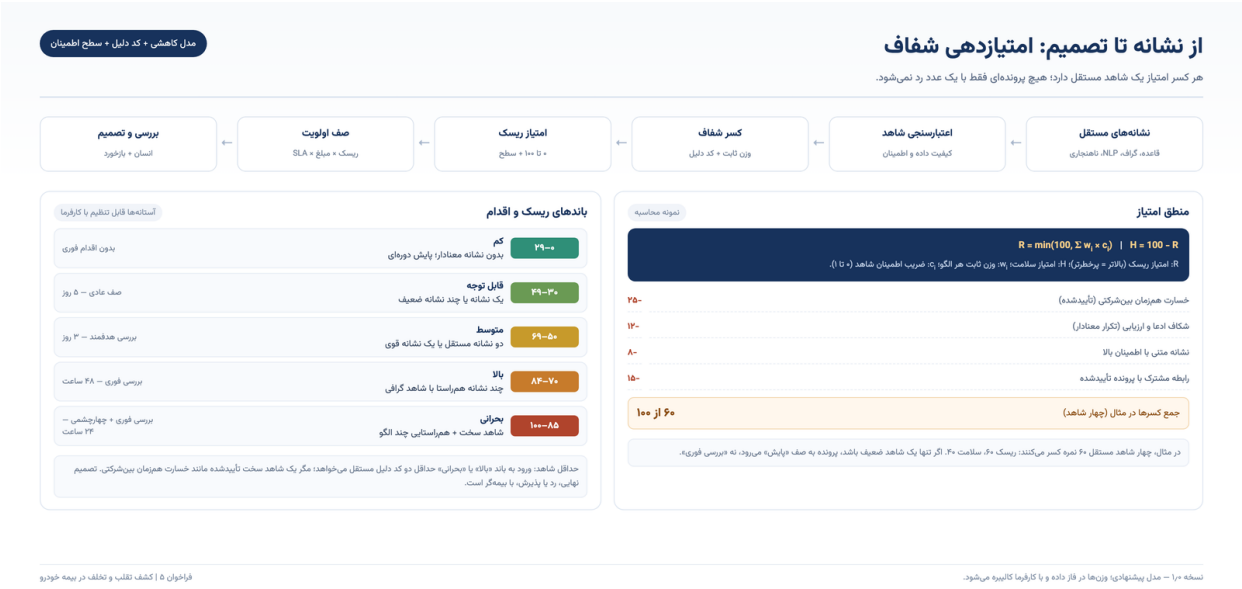
۱۲ امتیازدهی، اولویت‌بندی و جریان کار

امتیاز، زبان مشترک کارشناس، مدیر و ناظر است. اما یک عدد بدون دلیل، بی‌فایده و خطرناک است. این فصل، منطق امتیاز، باندهای ریسک، اولویت‌بندی، جریان کار و توضیح‌پذیری را مشخص می‌کند.

۱۲.۱ منطق امتیاز: ریسک و سلامت

برای رفع ابهام معنایی، دو عدد نمایش داده می‌شود: **امتیاز ریسک (R)** که از ۰ آغاز و با شواهد افزایش می‌یابد، و **امتیاز سلامت (H)** که مکمل آن است (`H = 100 − R`). مبنای اولویت‌بندی، امتیاز ریسک است؛ امتیاز سلامت برای ارتباط با بیمه‌گذار و گزارش مدیریتی استفاده می‌شود.

> **فرمول**
>
> $R = \min(100, \sum w_i \times c_i)$ و $H = 100 - R$. در این فرمول w_i وزن ثابت هر الگو و c_i ضریب اطمینان شاهد (۰ تا ۱) است. اگر اطمینان کمتر از ۰٫۵ باشد، شاهد فقط «اطلاعاتی» است و وارد امتیاز نمی‌شود.



شکل ۱۲-۱ — از نشانه تا تصمیم: منطق امتیاز و باندهای ریسک

۱۲.۲ نمونه محاسبه گام‌به‌گام

جدول ۱۲-۱ — نمونه محاسبه امتیاز

شاهد	کد دلیل	وزن	اطمینان	کسر	توضیح
خسارت هم‌زمان بین‌شرکتی	01-DUP	۲۵	۱٫۰	۲۵	تطبیق رخداد ۰٫۸۹
شکاف ادعا/پرداخت	01-GAP	۱۲	۰٫۸	۹٫۶	دو پرونده از سه
نشانه متنی	01-NLP	۱۰	۰٫۸	۸	دو جمله شاهد
رابطه مشترک	01-REL	۱۵	۰٫۷	۱۰٫۵	تلفن مشترک با پرونده تأییدشده
جمع	—	—	—	۵۳٫۱	ریسک ۵۳، سلامت ۴۷

در این مثال، چهار شاهد مستقل به امتیاز ریسک ۵۳ می‌رسند؛ پرونده در باند «متوسط» قرار می‌گیرد و به بررسی هدفمند سه‌روزه ارجاع می‌شود. اگر شاهد «صحنه‌سازی تأییدشده» هم اضافه شود (۲۰ نمره)، امتیاز به ۷۳ می‌رسد و به باند «بالا» می‌رود. این شفافیت، همان چیزی است که کارشناس و ناظر می‌توانند درباره آن گفت‌وگو کنند.

۱۲.۳ باندهای ریسک و سطح خدمت

جدول ۱۲-۲ — باندهای ریسک و SLA

باند	امتیاز ریسک	اقدام	SLA	سطح بازبینی
کم	۰ تا ۲۹	پایش دوره‌ای	—	خودکار / نمونه‌ای
قابل توجه	۳۰ تا ۴۹	صف عادی	۵ روز	کارشناس
متوسط	۵۰ تا ۶۹	بررسی هدفمند	۳ روز	کارشناس ارشد
بالا	۷۰ تا ۸۴	بررسی فوری	۴۸ ساعت	کارشناس + سرپرست
بحرانی	۸۵ تا ۱۰۰	بررسی فوری + چهارچشمی	۲۴ ساعت	کمیته بررسی

۱۲.۴ اولویت‌بندی فراتر از امتیاز

امتیاز تنها عامل نیست. یک پرونده با ریسک ۶۰ و مبلغ ۵۰۰ میلیون، ممکن است از پرونده با ریسک ۸۰ و مبلغ ۱۰ میلیون مهم‌تر باشد. اولویت از ترکیب **ریسک**، **مبلغ**، **عدم‌قطعیت** و **SLA** ساخته می‌شود. این ترکیب، ظرفیت محدود کارشناسان را به بهترین شکل تخصیص می‌دهد.

جدول ۱۲-۳ — عوامل اولویت‌بندی

عامل	وزن پیشنهادی	دلیل
امتیاز ریسک	۴۰٪	شدت شواهد
مبلغ در معرض خطر	۳۰٪	اثر مالی
عدم‌قطعیت	۱۵٪	نیاز به بررسی انسانی
SLA / قدمت پرونده	۱۵٪	الزام زمانی

۱۲.۵ جریان کار کارشناسی

جدول ۱۲-۴ — جریان کار

گام	فعالیت	نقش	خروجی
۱. هشدار	تولید امتیاز و کد دلیل	موتور	پرونده در صف
۲. غربال	بررسی اولیه و حذف هشدارهای تکراری	سرپرست صف	ارجاع یا بستن
۳. تخصیص	انتساب بر اساس تخصص و ظرفیت	مدیر	کارشناس
۴. بررسی	مدارک، گراف، متن، تماس	کارشناس	یادداشت و شاهد
۵. تصمیم	تأیید، رد، درخواست مدرک	کارشناس/کمیته	تصمیم مستند
۶. بازخورد	ثبت برچسب و دلیل	کارشناس	داده آموزش
۷. بستن	به‌روزرسانی وضعیت	سیستم	پرونده بسته

۱۲.۶ میز کار کارشناسی

میز کار باید در کمتر از ۳۰ ثانیه به کارشناس بگوید «چرا این پرونده مهم است». صفحه پرونده شامل این بخش‌هاست: **نوار امتیاز** (ریسک، سلامت، باند)، **کدهای دلیل** با شاهد، **مسیر گراف** (حداقل زیرگراف)، **خط زمانی** پرونده، **اسناد و متن‌های مرتبط**، **پرونده‌های مشابه**، **پرسش‌های پیشنهادی**، و **پنل تصمیم** با گزینه‌های تأیید/رد/درخواست مدرک. هر اقدام، در لاگ ممیزی ثبت می‌شود.

طراحی عامدانه ساده است: کارشناس نباید برای فهم هشدار، آموزش مدل ببیند. هر کد دلیل، یک جمله توضیح و یک شاهد قابل مشاهده دارد. اگر شاهد قابل نمایش نباشد (به‌دلیل محدودیت افشا)، به‌جای آن «نوع شاهد» و «سطح اطمینان» نمایش داده می‌شود.

۱۲.۷ توضیح‌پذیری در چهار سطح

جدول ۱۲-۵ — سطوح توضیح‌پذیری

سطح	مخاطب	محتوا
سطح ۱ — عدد	مدیر / داشبورد	امتیاز ریسک، باند، روند
سطح ۲ — کد دلیل	کارشناس	شناسه قاعده، توضیح کوتاه، وزن
سطح ۳ — شاهد	کارشناس / ناظر	مسیر گراف، جمله متن، نسبت مبلغ
سطح ۴ — بازتولید	ممیز / ناظر	نسخه قواعد، داده ورودی، محاسبه

۱۲.۸ اعتراض، بازبینی و مسئولیت

شرکت بیمه و بیمه‌گذار باید بتوانند کدهای دلیل را ببینند و بابت هر کد اعتراض ثبت کنند. سامانه، پاسخ و نتیجه بازبینی را در پرونده نگه می‌دارد. برای پرونده‌های باند «بحرانی»، بازبینی چهارچشمی الزامی است. تصمیم نهایی و مسئولیت آن با بیمه‌گر است؛ سامانه فقط شواهد و اولویت را فراهم می‌کند. این تفکیک، هم از کارشناس محافظت می‌کند و هم از ادعای «تصمیم ماشین» جلوگیری می‌کند.

۱۲.۹ داشبوردها

جدول ۱۲-۶ — داشبوردها

داشبورد	کاربر	شاخص‌ها
عملیاتی	سرپرست صف	حجم هشدار، زمان بررسی، نرخ تأیید، عقب‌افتادگی
مدیریتی	مدیر خسارت	روند نشت، ارزش جلوگیری‌شده، ROI، مقایسه دوره
ناظر	بیمه مرکزی	نمای تجمیعی، نسخه قواعد، ممیزی، رخداد
کیفیت داده	تیم داده	کامل‌بودن، اعتبار، تأخیر، خطای اتصال

۱۲.۱۰ تصمیم انسانی و محدودیت‌های اخلاقی

سه قاعده غیرقابل‌مذاکره: **هیچ رد خودکاری** بر اساس امتیاز وجود ندارد؛ **هیچ پرونده‌ای** فقط با یک نشانه ضعیف به باند بالا نمی‌رود؛ و **هیچ ویژگی حساسی** (قومیت، مذهب، وضعیت سلامت) در مدل استفاده نمی‌شود. متغیرهای مکانی و سنی فقط در بازه‌های تجمیعی و با توجیه ریسک به کار می‌روند. سوگیری مدل به‌صورت فصلی بررسی می‌شود.

۱۳ اعتبارسنجی، آزمون و طراحی پایلوت

یک ادعای فنی بدون آزمون، یک فرضیه است. این فصل، چارچوب ارزیابی، داده و برچسب، روش اعتبارسنجی، پایلوت سایه، آزمون‌ها، معیار پذیرش و طراحی آزمایشگاه داده را مشخص می‌کند.

۱۳.۱ چارچوب ارزیابی

جدول ۱۳-۱ — شاخص‌های ارزیابی

بُعد	شاخص	تعریف	هدف پیشنهادی
فنی	50@Precision	نسبت هشدارهای تأییدشده در ۵۰ هشدار اول	≥ ۷۵٪
فنی	Recall	نسبت پرونده‌های تقلب تأییدشده که هشدار گرفته‌اند	≥ ۵۵٪
فنی	AUC-PR	مساحت زیر منحنی precision-recall	بهبود ≥ ۲۰٪ نسبت به خط پایه
عملیاتی	زمان بررسی	میانگین زمان تا تصمیم	کاهش ≥ ۳۰٪
عملیاتی	نرخ مثبت کاذب	هشدارهای ردشده / کل هشدارهای بررسی‌شده	≥ ۲۵٪
کسب‌وکار	ارزش جلوگیری‌شده	طبق تعریف فصل ۳	≥ ۱۰ میلیارد تومان در پایلوت
اعتماد	پذیرش توضیح	نسبت کارشناسانی که توضیح را کافی دانستند	≥ ۸۰٪

۱۳.۲ داده و برچسب

اعتبارسنجی روی داده تاریخی ۲۴ تا ۳۶ ماه انجام می‌شود. برچسب‌ها از سه منبع می‌آیند: پرونده‌های تأییدشده تقلب، تصمیم‌های رد/کاهش با کد علت، و بازبینی نمونه‌ای کارشناسان. برچسب «مشکوک» به‌عنوان تقلب قطعی تلقی نمی‌شود. برای جلوگیری از نشت اطلاعات، تقسیم داده **زمانی** است: آموزش روی دوره قدیم، آزمون روی دوره جدید. پرونده‌هایی که پس از تاریخ آزمون ایجاد شده‌اند، در آموزش استفاده نمی‌شوند.

۱۳.۳ روش اعتبارسنجی

جدول ۱۳-۲ — روش‌های اعتبارسنجی

روش	توضیح	خروجی
گذشته‌نگر	اجرای مدل روی پرونده‌های تاریخی با برچسب	AUC-PR/Recall/Precision
تقسیم زمانی	آموزش دوره قدیم، آزمون دوره جدید	عملکرد واقعی‌تر
بازنمونه‌گیری	Bootstrap برای بازه اطمینان	CI 95٪
تحلیل خطا	بررسی هشدارهای نادرست	الگوهای شکست
مقایسه با خط پایه	قواعد فعلی بیمه‌گر	بهبود نسبی
پایلوت سایه	اجرای موازی بدون تأثیر بر تصمیم	پذیرش کارشناس

۱۳.۴ پایلوت سایه

در پایلوت سایه، سامانه به‌صورت موازی با فرآیند فعلی اجرا می‌شود، اما هیچ تصمیمی را تغییر نمی‌دهد. کارشناسان هشدارها را می‌بینند و نظر می‌دهند؛ نتیجه با تصمیم واقعی مقایسه می‌شود. این روش، هم دقت را می‌سنجد و هم تجربه کاربری را بدون ریسک عملیاتی ارزیابی می‌کند. مدت پیشنهادی: ۴ هفته، با حداقل ۲۰۰ پرونده بررسی‌شده.

۱۳.۵ آزمون‌ها

جدول ۱۳-۳ — آزمون‌ها

آزمون	دامنه	معیار خروج
واحد	قواعد، توکن، API	پوشش ≥ ۸۰٪ کد حیاتی
یکپارچه	اتصال بیمه‌گر به هاب	۱۰۰٪ سناریوهای بحرانی
کارایی	استعلام و batch	p95 > ۲s، batch در پنجره
امنیت	نفوذ، دسترسی، رمزنگاری	بدون آسیب‌پذیری بحرانی
UAT	میز کار و جریان تصمیم	تأیید کارشناس
بازیابی	RPO/RTO	بازیابی در زمان هدف

۱۳.۶ معیارهای پذیرش پایلوت

معیارهای زیر **پیشنهادی** هستند و پیش از شروع با کارفرما نهایی می‌شوند. هدف، ایجاد یک توافق شفاف است، نه تحمیل عدد.

جدول ۱۳-۴ — معیارهای پذیرش پیشنهادی

معیار	آستانه	نحوه سنجش
پوشش داده	≥ ۹۰٪	نسبت رکوردهای کامل و معتبر
Precision@50	≥ ۷۵٪	برچسب تأییدشده
Recall	≥ ۵۵٪	پرونده‌های تأییدشده
مثبت کاذب	≤ ۲۵٪	نمونه بررسی‌شده
زمان پاسخ	p95 < ۲s	آزمون بار
دسترس‌پذیری	≥ ۹۹٫۵٪	پایش
کاهش زمان بررسی	≥ ۳۰٪	مقایسه با خط پایه
رضایت کارشناس	≥ ۴ از ۵	پرسشنامه
پرونده تأییدشده	≥ ۲۰	گزارش کمیته

۱۳.۷ طراحی آزمایشگاه داده

آزمایشگاه داده، محیط کنترل‌شده‌ای است که بیمه مرکزی در اختیار تیم‌های منتخب قرار می‌دهد. ما سه الزام برای این محیط پیشنهاد می‌کنیم: **دسترسی به داده بی‌نام/شبه‌واقعی** با قرارداد استفاده؛ **جداسازی محیط** از شبکه تولید بیمه‌گران؛ و **پایش فعالیت** برای جلوگیری از خروج داده. محیط باید شامل یک نمونه داده مرجع، ابزار ارزیابی و امکان اجرای پایلوت سایه باشد.

۱۳.۸ گزارش پایلوت

گزارش نهایی شامل این بخش‌هاست: خلاصه مدیریتی، توصیف داده، روش‌ها، نتایج فنی (PR/recall/precision-AUC)، نتایج عملیاتی (زمان، حجم، پذیرش)، تحلیل خطا، ارزش اقتصادی، ریسک‌ها و توصیه توسعه. گزارش باید قابل بازتولید باشد: نسخه قواعد، نسخه مدل، تاریخ داده و پارامترها در پیوست ذکر می‌شوند.

۱۳.۹ تصمیم go-no/go

جدول ۱۳-۵ — چارچوب تصمیم

نتیجه	شرط	اقدام
Go	همه معیارهای بحرانی محقق شود	قرارداد تولید و برنامه توسعه
Conditional Go	یک یا دو معیار نزدیک به هدف	رفع نقص در ۴ تا ۶ هفته و بازآزمون
Go-No	شکست در داده، امنیت یا دقت پایه	توقف، تحلیل علت، بازطراحی

۱۴ امنیت، انطباق و حاکمیت داده

امنیت در این طرح، یک لایه اضافه نیست؛ شرط وجودی است. اگر بیمه‌گر و ناظر به کنترل دسترسی و عدم نشت اطمینان نداشته باشند، هیچ مدلی اجرا نمی‌شود. این فصل، چارچوب انطباق، طبقه‌بندی داده، کنترل‌ها، مدیریت دسترسی، رمزنگاری، ممیزی، پاسخ به حادثه و مسیر گواهینامه را مشخص می‌کند.

۱۴.۱ چارچوب انطباق

جدول ۱۴-۱ — چارچوب انطباق

مرجع / الزام	ارتباط با طرح	اقدام
قانون بیمه و آیین‌نامه‌های بیمه مرکزی	مجوز فعالیت، محرمانگی، نظارت	مشورت حقوقی و تأیید ناظر
قانون تجارت الکترونیک	حفاظت داده در تراکنش‌های الکترونیکی	قرارداد داده، رضایت/مبنا، امنیت
قانون جرایم رایانه‌ای	دسترسی غیرمجاز و تخریب داده	کنترل دسترسی، لاگ، آزمون نفوذ
الزامات افتا	امنیت زیرساخت و سامانه‌های حیاتی	ارزیابی امنیتی، مستندسازی
مصوبات شورای عالی فضای مجازی	حاکمیت داده و اقامت	استقرار داخلی، انطباق
قانون حفاظت از داده‌های شخصی (در صورت ابلاغ)	حقوق افراد، مبنای پردازش	DPIA، ثبت پردازش، به‌روزرسانی
قانون مبارزه با پولشویی	ذی‌نفع و تراکنش مشکوک	همکاری با واحد AML بیمه‌گر

این جدول، فهرست جامع قوانین نیست؛ نقشه ارتباط است. پیش از پایلوت، یک **یادداشت حقوقی** تهیه می‌شود که مبنای پردازش، نقش طرف‌ها، مدت نگهداری و شرایط افشا را تثبیت کند. در صورت تغییر مقررات، DPIA و قراردادها بازبینی می‌شوند.

۱۴.۲ طبقه‌بندی داده

جدول ۱۴-۲ — طبقه‌بندی داده

طبقه	نمونه	محل نگهداری	دسترسی	نگهداری
عمومی	گزارش تجمیعی، بنچمارک	داشبورد ناظر	همه اعضا	نامحدود
داخلی	مستندات فنی، نسخه قواعد	مخزن کد/سند	تیم فنی	۳ سال
محرمانه	داده توکن‌شده، ویژگی‌ها	هاب تحلیلی	نقش‌های مجاز	۳۶ ماه
محدود	شناسه خام در لحظه توکن	حافظه سرویس توکن	سرویس توکن	ذخیره نمی‌شود
حساس	کلید، اعتبارنامه، لاگ ممیزی	SIEM / HSM	نگهدارنده کلید/ممیز	طبق سیاست

۱۴.۳ کنترل‌های امنیتی

جدول ۱۴-۳ — کنترل‌های امنیتی

نوع	کنترل	هدف
پیشگیرانه	mTLS، OAuth2/OIDC، MFA، ABAC/RBAC	جلوگیری از دسترسی غیرمجاز
پیشگیرانه	رمزنگاری TLS 1.3 و AES-256 در حالت سکون	حفاظت داده در انتقال و ذخیره
پیشگیرانه	KMS/HSM، چرخش کلید، تفکیک وظایف	حفاظت کلید
پیشگیرانه	WAF، segmentation، حداقل پورت	کاهش سطح حمله
پیشگیرانه	امنیت زنجیره تأمین، SBOM، اسکن وابستگی	جلوگیری از آسیب‌پذیری
تشخیصی	SIEM، IPS/IDS، پایش رفتار	کشف نفوذ
تشخیصی	پایش استعلام انبوه و نشت	کشف سوءاستفاده
تشخیصی	اسکن آسیب‌پذیری دوره‌ای	شناسایی ضعف
اصلاحی	پاسخ به حادثه، بازیابی، وصله	کاهش اثر
اصلاحی	آزمون نفوذ سالانه	اعتبارسنجی کنترل

۱۴.۴ مدیریت دسترسی و ماتریس نقش‌ها

جدول ۱۴-۴ — ماتریس دسترسی

نقش	داده خام	داده توکن‌شده	امتیاز/شاهد	مدیریت قواعد	ممیزی
کارشناس بیمه‌گر	—	—	فقط پرونده‌های خود	—	—
تحلیل‌گر تقلب	—	تجمیعی	نمای تجمیعی	—	—
مدیر داده	—	بله	—	—	—
نگهدارنده کلید	—	—	—	—	—
مدیر سامانه	—	محدود	—	بله	—
ناظر / ممیز	—	—	گزارش	—	بله
DPO	—	—	گزارش	—	بله

تفکیک وظایف: نگهدارنده کلید به داده دسترسی ندارد؛ مدیر داده به کلید دسترسی ندارد؛ تحلیل‌گر به هویت دسترسی ندارد. دسترسی ممتاز با MFA، ثبت و بازبینی دوره‌ای همراه است. هر دسترسی موقت، تاریخ انقضا دارد.

۱۴.۵ رمزنگاری و مدیریت کلید

- **در انتقال:** TLS 1.3 برای همه APIها؛ SFTP با PGP برای فایل‌های دسته‌ای.
- **در سکون:** AES-256 برای پایگاه داده، شیء storage و پشتیبان.
- **کلید:** HSM یا KMS با تفکیک نقش؛ هیچ کلیدی در کد، فایل پیکربندی یا لاگ نیست.
- **چرخش:** کلید رمزنگاری سالانه؛ کلید توکن با نسخه‌بندی و حفظ کلید قدیم برای اتصال.
- **بازیابی:** رویه بازیابی کلید با تأیید چندنفره و ثبت ممیزی.

۱۴.۶ ثبت، ممیزی و عدم انکار

همه رخدادهای حساس ثبت می‌شوند: ورود، استعلام، تغییر قاعده، تغییر دسترسی، صدور امتیاز، تصمیم کارشناس و تغییر داده مرجع. لاگ‌ها در مخزن **تغییرناپذیر (WORM)** با مهر زمانی نگهداری می‌شوند. دسترسی به لاگ، خواندنی و فقط برای نقش ممیزی است. هر پاسخ API، شناسه درخواست یکتا دارد تا بتوان یک تصمیم را به ورودی و نسخه قواعد آن بازگرداند.

۱۴.۷ پاسخ به حادثه و تداوم کسب‌وکار

جدول ۱۴-۵ — پاسخ به حادثه

مرحله	فعالیت	مسئول	زمان
تشخیص	هشدار SIEM / گزارش کاربر	SOC	≤ ۱۵ دقیقه
مهار	قطع دسترسی، جداسازی	امنیت	≤ ۱ ساعت
ارزیابی	دامنه، داده، اثر	تیم حادثه	≤ ۴ ساعت
اطلاع‌رسانی	بیمه‌گر، ناظر، DPO	مدیریت	≤ ۲۴ ساعت
بازیابی	بازگردانی سرویس	عملیات	طبق RTO
درس‌آموخته	گزارش و اصلاح	همه	≤ ۷ روز

تداوم کسب‌وکار: RTO چهار ساعت و RPO یک ساعت برای پایلوت؛ در تولید، RTO یک ساعت و RPO ۱۵ دقیقه هدف‌گذاری می‌شود. پشتیبان‌گیری روزانه، رمزنگاری‌شده و خارج از سایت؛ آزمون بازیابی فصلی. سناریوی از کار افتادن هاب، شامل اجرای موقت در حالت خواندنی و صف محلی است.

۱۴.۸ حریم خصوصی و DPIA

DPIA پیش از پایلوت انجام و سالانه بازبینی می‌شود. خروجی آن شامل نقشه پردازش، ارزیابی ضرورت، ریسک‌ها، کنترل‌ها و تأیید DPO است. حقوق افراد از طریق بیمه‌گر پاسخ داده می‌شود؛ درخواست دسترسی، اصلاح یا اعتراض در یک بازه زمانی مشخص ثبت و پیگیری می‌شود. داده‌های خارج از دوره نگهداری، به‌صورت امن حذف و گواهی حذف صادر می‌شود.

۱۴.۹ مدیریت ریسک فروشنده و زنجیره تأمین

هر زیرپردازشگر پیش از قرارداد ارزیابی می‌شود: گواهینامه‌های امنیتی، سابقه رخداد، دسترسی به داده، محل استقرار و امکان خروج. قرارداد شامل حق ممیزی، الزام اطلاع‌رسانی رخداد، محدودیت زیرپردازشگر و حذف داده در پایان است. فهرست نرم‌افزارهای متن‌باز و نسخه‌ها (SBOM) نگهداری و آسیب‌پذیری‌ها ماهانه بررسی می‌شود.

۱۴.۱۰ مسیر گواهینامه‌ها

جدول ۱۴-۶ — مسیر گواهینامه‌ها

گواهینامه / ارزیابی	وضعیت هدف	زمان تقریبی	ارزش برای کارفرما
IEC 27001/ISO	در مسیر دریافت	۹ تا ۱۲ ماه	اعتماد به مدیریت امنیت
IEC 27701/ISO	پس از 27001	۳ تا ۶ ماه	مدیریت حریم خصوصی
ارزیابی افتا	پیش از تولید	۳ تا ۶ ماه	انطباق ملی
آزمون نفوذ مستقل	پیش از live-go	۴ تا ۶ هفته	اعتبار فنی

> **ادعای صادقانه**
>
> ما در این پیشنهاد مدعی گواهینامه‌ای که هنوز دریافت نشده نیستیم. تعهد ما، مسیر مشخص، مستندسازی و پذیرش ممیزی مستقل است. گواهینامه، نتیجه اجراست، نه پیش‌شرط ادعا.

۱۴.۱۱ حاکمیت داده بین‌بیمه‌ای

حاکمیت داده در این طرح سه سطح دارد: **سطح عضو** (مسئولیت داده خود)، **سطح نهاد امین** (نگهداری کلید، مجوز دسترسی، ممیزی) و **سطح ناظر** (سیاست‌گذاری، بازبینی، رسیدگی به شکایت). تغییر قواعد امتیازدهی نیازمند تأیید کمیته فنی و اطلاع ناظر است. رخداد امنیتی مهم، حداکثر ۲۴ ساعت به اعضا و ناظر اطلاع داده می‌شود.

۱۴.۱۲ انطباق با الزامات فراخوان

فراخوان، داده‌های بی‌نام و دسترسی کنترل‌شده در آزمایشگاه داده را الزامی کرده است. طراحی ما این الزام را نه به‌عنوان یک گزینه، بلکه به‌عنوان **معماری پایه** پذیرفته است: توکن‌سازی در نقطه ورود، عدم ذخیره شناسه خام، حداقل‌سازی، کنترل استعلام و ممیزی. جزئیات فنی در فصل ۹ و ۱۰ آمده است.

۱۵ برنامه اجرا و تیم

برنامه ۱۲ هفته‌ای عامدانه فشرده اما واقع‌گرایانه است: چهار دروازه تصمیم، جریان‌های کاری موازی و وابستگی‌های روشن به کارفرما. این فصل، زمان‌بندی، تیم، تجربه و مدل همکاری را مشخص می‌کند.

۱۵.۱ اصول برنامه‌ریزی

- **شروع کوچک:** دو بیمه‌گر، یک خط محصول، یک محیط.
- **تحویل تدریجی:** هر دو هفته یک خروجی قابل مشاهده.
- **موازی‌سازی:** داده، گراف، NLP و امنیت همزمان اما با وابستگی مشخص.
- **دروازه تصمیم:** ادامه یا توقف در چهار نقطه، نه در پایان پروژه.
- **مشارکت کارفرما:** دسترسی داده، تأیید مدل و معرفی کارشناس در زمان مقرر.

۱۵.۲ برنامه ۱۲ هفته‌ای



شکل ۱۵-۱ — برنامه ۱۲ هفته‌ای و چهار دروازه تصمیم

جدول ۱۵-۱ — برنامه هفتگی

هفته	جریان کار	فعالیت‌ها	خروجی
۱	استقرار	جلسه شروع، نقش‌ها، دسترسی، تقویم	منشور پروژه
۲	داده	نقشه داده، نمونه، توافق امنیتی	قرارداد داده و نمونه
۳	داده	بارگذاری تاریخی، کیفیت داده	گزارش DQ
۴	داده	توکن‌سازی، مدل کانونیکال	لایه داده آماده
۵	گراف	ساخت گراف، تطبیق رخداد	گراف پایه
۶	قواعد	کاتالوگ قواعد، موتور امتیاز	نسخه ۱ قواعد
۷	NLP	خط لوله متن، برچسب نمونه	مدل NLP اولیه
۸	میز کار	صف، صفحه پرونده، بازخورد	میز کار قابل استفاده
۹	یکپارچه	API، امنیت، UAT	اتصال کامل
۱۰	آزمون	اعتبارسنجی گذشته‌نگر، رفع نقص	گزارش دقت
۱۱	پایلوت سایه	اجرای موازی، جمع‌آوری بازخورد	نتایج پایلوت
۱۲	تحویل	گزارش، آموزش، تصمیم	گزارش نهایی و توصیه

۱۵.۳ دروازه‌های تصمیم

جدول ۱۵-۲ — دروازه‌های تصمیم

دروازه	زمان	معیار عبور	تصمیم
G1	پایان هفته ۲	قرارداد داده، نمونه، محیط امن	شروع توسعه
G2	پایان هفته ۵	پوشش داده ≥ ۹۰٪، گراف پایه	ادامه یا اصلاح داده
G3	پایان هفته ۹	موتور کامل، میز کار، اتصال	شروع پایلوت
G4	پایان هفته ۱۲	معیارهای پذیرش	توسعه یا توقف

۱۵.۴ تیم و نقش‌ها

جدول ۱۵-۳ — تیم پیشنهادی

نقش	نام	تخصیص	مسئولیت	صلاحیت
سرپرست محصول	نام	۱۰۰٪	راهبرد، کارفرما، پذیرش	محصول داده/بیمه
سرپرست فنی	نام	۱۰۰٪	معماری، کیفیت، تحویل	معماری داده و نرم‌افزار
مهندس داده ارشد	نام	۱۰۰٪	مدل داده، ETL، کیفیت	Airflow/Python/SQL
مهندس بک‌اند	نام	۱۰۰٪	API، سرویس، امنیت	Postgres/FastAPI
متخصص NLP/ML	نام	۸۰٪	مدل، اعتبارسنجی، NLP	ParsBERT/Python
تحلیل‌گر گراف	نام	۶۰٪	گراف، الگو،Query	NetworkX/Neo4j
کارشناس خسارت	نام	۵۰٪	قواعد، اعتبارسنجی، آموزش	بیمه خودرو
متخصص امنیت/انطباق	نام	۵۰٪	DPIA، دسترسی، آزمون	امنیت/حقوق
طراح UX	نام	۴۰٪	میز کار، داشبورد	طراحی محصول
PMO	نام	۵۰٪	زمان‌بندی، ریسک، گزارش	مدیریت پروژه

در پایلوت، برخی نقش‌ها به‌صورت پاره‌وقت یا مشاور تأمین می‌شوند؛ تخصیص نهایی پس از توافق دامنه و تقویم با کارفرما تثبیت می‌شود.

۱۵.۵ تجربه مرتبط

تیم ما پیش از این، **موتور بی‌نام‌سازی** را در سطح کد مهندسی ساخته و اجرا کرده است: خواندن فیلدهای پایگاه داده، حذف ستون‌های هویتی و تبدیل شناسه‌ها به توکن یک‌طرفه. این کد، نقطه شروع لایه اعتماد است و در این پروژه به یک محصول صنعتی با HSM، نسخه‌بندی و ممیزی تبدیل می‌شود. همچنین تجربه اجرای پروژه‌های داده در مقیاس ملی/منطقه‌ای، زیرساخت داده و تحلیل مکانی در تیم وجود دارد.

جدول ۱۵-۴ — شواهد توان تیم (نیازمند تکمیل)

قابلیت	شاهد	ارتباط با طرح
موتور بی‌نام‌سازی	کد مهندسی موجود	لایه توکن‌سازی
پروژه‌های داده مکانی	نام پروژه/مرجع	مدل داده و گراف
زیرساخت داده	نام پروژه/مرجع	انبار داده و ETL
تحلیل و مدل	نمونه	امتیازدهی و NLP
امنیت و انطباق	نمونه	DPIA و کنترل

> **برای تکمیل پیش از ارسال**
>
> سه مرجع مشتری، یک نمونه خروجی از موتور بی‌نام‌سازی (بدون داده واقعی) و خلاصه سه پروژه مرتبط را آماده کنید. ارزیاب به شاهد عینی نیاز دارد، نه توصیف کلی.

۱۵.۶ مدل همکاری با کارفرما

جدول ۱۵-۵ — مکانیزم همکاری

مکانیزم	فراوانی	شرکت‌کنندگان	خروجی
کمیته راهبری	دوهفتگی	کارفرما، ناظر، سرپرست‌ها	تصمیم و رفع مانع
جلسه فنی	هفتگی	تیم فنی دو طرف	پیشرفت و وابستگی
نمایش محصول	دوهفتگی	کارشناسان	بازخورد
گزارش وضعیت	هفتگی	PMO	ریسک، زمان، کیفیت

۱۵.۷ مدیریت تغییر و آموزش

پذیرش سامانه به آموزش وابسته است. سه برنامه پیشنهاد می‌شود: **کارگاه کارشناسان** (۲ جلسه، شامل مطالعه پرونده و تصمیم)، **راهنمای تفسیر کد دلیل** (یک صفحه برای هر الگو)، و **آموزش مدیران** (داشبورد و خواندن گزارش). بازخورد کاربران در دو هفته اول پایلوت، به اصلاح میز کار و متن توضیح‌ها منجر می‌شود.

۱۵.۸ پس از پایلوت: مسیر توسعه

جدول ۱۵-۶ — مسیر توسعه

دوره	هدف	اقدام
۰ تا ۳ ماه	تثبیت پایلوت	رفع نقص، کالیبراسیون، مستندسازی
۳ تا ۹ ماه	تولید	اتصال ۶ تا ۸ بیمه‌گر، SLA، پشتیبانی
۹ تا ۱۸ ماه	شبکه	بنچمارک، خطوط دیگر، API ecosystem
۱۸ ماه+	توسعه فناوری	یادگیری فدرال، PSI، حریم خصوصی تفاضلی

۱۶ ریسک‌ها، شاخص‌ها و ارزش اثباتی

هیچ طرحی بدون ریسک نیست؛ طرح معتبر، ریسک‌ها را می‌شناسد و پاسخ دارد. این فصل، ریسک‌های کلیدی، شاخص‌های موفقیت، سناریوهای شکست و ملاحظات اخلاقی را مشخص می‌کند.

۱۶.۱ ریسک‌های کلیدی و پاسخ‌ها

جدول ۱۶-۱ — ریسک‌ها

ریسک	احتمال	اثر	پاسخ	مالک
تأخیر دسترسی به داده	بالا	بالا	شروع با نمونه، توافق زودهنگام، دروازه G1	PMO
کیفیت پایین داده	بالا	متوسط	DQ gate، قرنطینه، برنامه اصلاح	تیم داده
عدم مشارکت بیمه‌گر دوم	متوسط	بالا	نقش ناظر، مشوق متقابل، جایگزین	مدیریت
مثبت کاذب بالا	متوسط	بالا	آستانه اطمینان، بازبینی انسانی، کالیبراسیون	ML
نشت داده	کم	بحرانی	توکن، HSM، آزمون نفوذ، DPIA	امنیت
ابهام حقوقی	متوسط	بالا	یادداشت حقوقی، پایلوت، تأیید ناظر	حقوقی
رانش مدل	متوسط	متوسط	پایش، بازآموزی، نسخه‌بندی	ML
وابستگی به فرد کلیدی	متوسط	متوسط	مستندسازی، جانشین، جفت‌کاری	PMO
پیچیدگی اتصال	متوسط	متوسط	batch اول، کانکتور استاندارد	فنی
اختلاف بر سر انتساب نتیجه	متوسط	متوسط	فرمول شفاف، کمیته مشترک، نمونه‌گیری	مدیریت

۱۶.۲ شاخص‌های موفقیت

جدول ۱۶-۲ — شاخص‌های موفقیت

بُعد	شاخص	هدف	فراوانی
فنی	50@Precision	≥ ۷۵٪	ماهانه
فنی	Recall	≥ ۵۵٪	ماهانه
فنی	نرخ مثبت کاذب	≤ ۲۵٪	ماهانه
عملیاتی	زمان بررسی	−۳۰٪	هفتگی
عملیاتی	دسترس‌پذیری	≥ ۹۹٫۵٪	ماهانه
کسب‌وکار	ارزش جلوگیری‌شده	≥ ۱۰ میلیارد تومان	فصلی
پذیرش	کاربران فعال	≥ ۸۰٪ کارشناسان	ماهانه
اعتماد	پذیرش توضیح	≥ ۸۰٪	فصلی
انطباق	یافته بحرانی ممیزی	صفر	سالانه

۱۶.۳ ارزش اثباتی برای ارزیاب

- **کد موجود:** موتور بی‌نام‌سازی از قبل ساخته شده؛ ادعای صفر تا صد نیست.
- **پایلوت کوچک:** ۱۲ هفته، دو بیمه‌گر، معیار پذیرش روشن.
- **معماری قابل اجرا:** API + batch، بدون جایگزینی سیستم بیمه‌گر.
- **توضیح‌پذیری:** کد دلیل، شاهد، نسخه قواعد؛ بدون جعبه سیاه.
- **مدل مالی هم‌راستا:** پرداخت مبتنی بر نتیجه، با تعریف قابل حسابرسی.
- **تیم چندتخصصی:** محصول، داده، ML، امنیت و کارشناس بیمه.

۱۶.۴ سناریوهای شکست و پاسخ

جدول ۱۶-۳ — سناریوهای شکست

سناریو	نشانه	پاسخ
داده کافی نیست	پوشش < ۷۰٪ در هفته ۴	توقف توسعه، اصلاح داده، جایگزین منبع
برچسب تقلب کم است	کمتر از ۵۰ پرونده تأییدشده	برچسب ضعیف + نمونه‌گیری + شبیه‌سازی
بیمه‌گر دوم وارد نمی‌شود	عدم توافق تا هفته ۳	پایلوت تک‌شرکتی + ارزش داخلی
مثبت کاذب بالاست	< ۴۰٪ در هفته ۱۰	کاهش وزن، آستانه بالاتر، بازبینی
ناظر ایراد حقوقی می‌گیرد	نامه در هفته ۲	توقف، DPIA، بازطراحی افشا
کارشناسان استفاده نمی‌کنند	پذیرش < ۵۰٪	آموزش، ساده‌سازی UI، بازخورد

۱۶.۵ ملاحظات اخلاقی

سه اصل اخلاقی بر طراحی حاکم است: **عدم اتهام بدون شاهد** (هیچ برچسبی بدون شاهد قابل بررسی صادر نمی‌شود)، **عدم تبعیض** (ویژگی‌های حساس استفاده نمی‌شوند و سوگیری فصلی بررسی می‌شود)، و **حق اعتراض** (بیمه‌گذار و بیمه‌گر می‌توانند تصمیم را بازبینی کنند). کمیته اخلاق با حضور نماینده بیمه‌گذار، سیاست‌های استفاده و موارد حاشیه‌ای را بررسی می‌کند.

۱۶.۶ محدودیت‌های تحلیل

- **مشاهده‌ای، نه علّی:** همبستگی‌ها نشانه هستند، نه اثبات علت.
- **سوگیری برچسب:** پرونده‌های بررسی‌شده، نماینده کل پرونده‌ها نیستند.
- **نمونه کوچک:** پایلوت دو بیمه‌گر، تعمیم‌پذیری محدود دارد.
- **تغییر رفتار:** با اعلام سامانه، ممکن است الگوی تقلب تغییر کند.
- **کیفیت داده:** نتیجه، به کیفیت داده منبع وابسته است.

۱۷ نتیجه‌گیری و درخواست

این طرح یک مزیت انحصاری نمی‌فروشد؛ یک تعارض منافع را حل می‌کند. مسئله از سال‌ها پیش روشن بوده و راه‌حل آن در دو عبارت خلاصه می‌شود: **بی‌نام‌سازی در لایه داده** و **امتیازدهی توضیح‌پذیر در لایه تحلیل**. شرکت بیمه، داده مشتری خود را از دست نمی‌دهد؛ در عوض دانش تجمیعی صنعت را به‌دست می‌آورد و ناظر نیز یک سامانه قابل ممیزی دریافت می‌کند.

۱۷.۱ جمع‌بندی استدلال

تقلب خودرو در مرز بین شرکت‌ها زندگی می‌کند. هر بیمه‌گر تنها یک قطعه را می‌بیند و به‌دلیل ترس تجاری و الزامات حقوقی، داده خام را به اشتراک نمی‌گذارد. راه‌حل، تجمیع داده نیست؛ ایجاد یک **شبکه شواهد** است که در آن داده هویتی در مبدأ توکن می‌شود، رخدادها با تطبیق دقیق و فازی به هم وصل می‌شوند، گراف رابطه‌ها را آشکار می‌کند، متن گزارش با NLP بررسی می‌شود، و امتیاز با کد دلیل و شاهد به کارشناس می‌رسد. این معماری، مانع تجاری را به‌صورت فنی رفع می‌کند و بار حقوقی را قابل مدیریت می‌کند.

اجرای ۱۲ هفته‌ای، عامدانه کوچک است: دو بیمه‌گر، یک خط محصول، یک محیط کنترل‌شده. هدف، اثبات سه چیز است: **می‌توان بدون افشای داده، الگوهای بین‌شرکتی را دید**؛ **می‌توان با شواهد محدود، دقت قابل قبول ساخت**؛ و **می‌توان مدل مالی را با نتیجه هم‌راستا کرد**. اگر این سه اثبات شوند، مسیر تولید و شبکه باز است.

۱۷.۲ درخواست ما

جدول ۱۷-۱ — درخواست‌های مشخص

درخواست	دلیل	زمان
دسترسی به آزمایشگاه داده	محیط کنترل‌شده برای توسعه و ارزیابی	هفته ۱
معرفی دو بیمه‌گر پایلوت	داده واقعی و بازخورد کارشناس	هفته ۱
حمایت دبیرخانه برای چارچوب حقوقی	تثبیت مبنای تبادل داده	هفته ۱ تا ۲
جلسه فنی با تیم داده بیمه‌گران	نقشه‌برداری و اتصال	هفته ۲
۱۲ هفته زمان و دسترسی به کارشناسان	اثبات عملی	هفته ۱ تا ۱۲

۱۷.۳ گام‌های بعدی

۱) **جلسه معرفی ۴۵ دقیقه‌ای:** ارائه دموی مفهومی و پاسخ به پرسش‌های ارزیاب.

۲) **کارگاه داده ۲ ساعته:** بررسی نمونه داده و نقشه اتصال با تیم فنی بیمه‌گر.

۳) **توافق‌نامه پایلوت:** دامنه، معیار پذیرش، مسئولیت‌ها و تقویم.

۴) **شروع هفته ۱:** استقرار، دسترسی و جلسه شروع.

۱۷.۴ تماس

برای هماهنگی جلسه فنی یا دریافت نسخه اجرایی، با تیم ما در تماس باشید:

جدول ۲-۱۷ — اطلاعات تماس (نیازمند تکمیل)

نقش	نام	تماس
سرپرست محصول و کسب‌وکار	نام	ایمیل | تلفن
سرپرست فنی و داده	نام	ایمیل | تلفن
هماهنگی پروژه	نام	ایمیل

> **جمله پایانی**
>
> ما نمی‌خواهیم داده رقبا را ببینیم؛ می‌خواهیم کاری کنیم که تقلب، در شبکه، دیده شود. این تفاوت، همان چیزی است که همکاری بین‌شرکتی را ممکن می‌کند.

پیوست‌ها

این پیوست‌ها برای ارزیاب فنی و تیم اجرایی طراحی شده‌اند: واژه‌نامه، نمونه API، کاتالوگ قواعد، چک‌لیست داده، فرم ارزیابی پایلوت، فهرست اطلاعات مورد نیاز و مراجع.

پیوست الف واژه‌نامه

جدول پ-۱ — واژه‌نامه

اصطلاح	معادل	تعریف
شبه‌بی‌نام‌سازی	Pseudonymization	جایگزینی شناسه با توکن؛ داده همچنان قابل پیوند است
توکن	Token	شناسه جایگزین تولیدشده با تابع کلیددار
HMAC	based Message Authentication Code-Hash	تابع درهم‌سازی کلیددار برای تولید توکن
HSM	Hardware Security Module	ماژول سخت‌افزاری نگهداری کلید
KMS	Key Management Service	سرویس مدیریت چرخه عمر کلید
گراف دانش	Knowledge Graph	مدل گره/یال برای نمایش رابطه موجودیت‌ها
تحلیل گراف	Graph Analytics	الگوریتم‌های خوشه، مسیر و مرکزیت
NLP	Natural Language Processing	پردازش زبان طبیعی
ParsBERT	Persian BERT	مدل زبانی فارسی مبتنی بر ترنسفورمر
XAI	Explainable AI	هوش مصنوعی توضیح‌پذیر
امتیاز ریسک	Risk Score	عدد ۰ تا ۱۰۰؛ بالاتر = پرخطرتر
امتیاز سلامت	Health Score	۱۰۰ منهای امتیاز ریسک
کد دلیل	Reason Code	شناسه قاعده/شاهد سازنده امتیاز
باند ریسک	Risk Band	دسته‌بندی امتیاز برای تعیین SLA
DPIA	Data Protection Impact Assessment	ارزیابی اثر حریم خصوصی
DPO	Data Protection Officer	افسر حریم خصوصی
RBAC	Based Access Control-Role	کنترل دسترسی نقش‌محور
ABAC	Based Access Control-Attribute	کنترل دسترسی ویژگی‌محور
SIEM	Security Information and Event Management	مدیریت رویداد امنیتی
RPO/RTO	Point Objective/Recovery Time	هدف زمان/نقطه بازیابی
SLA	Service Level Agreement	توافق سطح خدمت
Batch	پردازش دسته‌ای	انتقال داده در بازه‌های زمانی
API	Application Programming Interface	رابط برنامه‌نویسی
mTLS	Mutual TLS	احراز هویت دوطرفه در TLS
Idempotency	تکرارپذیری	اثر یکسان در تکرار درخواست
k-آنونیمیتی	Anonymity-k	حداقل k رکورد مشابه برای جلوگیری از بازشناسی
حریم خصوصی تفاضلی	Differential Privacy	افزودن نویز کنترل‌شده به گزارش تجمیعی
PSI	Private Set Intersection	اشتراک مجموعه بدون افشای عناصر

پیوست ب نمونه API

نمونه زیر برای نمایش قرارداد است؛ همه شناسه‌ها ساختگی و بدون داده واقعی هستند. پاسخ، هیچ فیلد هویتی یا مالی دقیق ندارد.

درخواست امتیاز

```
POST /v1/score
Authorization: Bearer <token>
Content-Type: application/json
Idempotency-Key: 7f3c...

{
  "request_id": "REQ-1405-000123",
  "insurer_code": "INS-A",
  "claim_token": "CLM-9F2A...",
  "subject_token": "SUB-3B7C...",
  "vehicle_token": "VEH-1D4E...",
  "event": {
    "date": "2026-07-14",
    "province_code": "23",
    "type": "COLLISION",
    "injury": false
  },
  "claim": {
    "claimed_amount_band": "400-500M",
    "coverage": ["BODY"],
    "report_date": "2026-07-16"
  },
  "purpose_code": "FRAUD_DETECTION",
  "rules_version": "1.4.2"
}
```

پاسخ

```
HTTP/1.1 200 OK
Content-Type: application/json

{
  "request_id": "REQ-1405-000123",
  "risk_score": 72,
  "health_score": 28,
  "band": "HIGH",
  "badges": ["MULTI_INSURER_DUPLICATE", "RELATION_CLUSTER"],
  "reason_codes": [
    {"code": "DUP-01", "weight": 25, "confidence": 0.89, "evidence": "EVENT_MATCH"},
    {"code": "GAP-01", "weight": 12, "confidence": 0.80, "evidence": "AMOUNT_RATIO"},
    {"code": "REL-01", "weight": 15, "confidence": 0.70, "evidence": "GRAPH_PATH"}
  ],
  "rules_version": "1.4.2",
  "model_version": "nlp-0.9.1",
  "sla": "48h",
  "explanation": "تطبیق رخداد با پرونده دیگر و رابطه مشترک تلفن؛ دو شاهد مستقل."
}
```

کدهای خطا

جدول پ-۲ — کدهای خطا

کد	معنا	اقدام
400	طرح درخواست نامعتبر	اصلاح فیلد و ارسال مجدد
401	احراز هویت ناموفق	بررسی توکن/mTLS
403	هدف یا نقش مجاز نیست	بررسی مجوز
404	پرونده یافت نشد	بررسی شناسه
429	سقف نرخ	تلاش مجدد با backoff
500	خطای داخلی	ثبت و پیگیری
503	سرویس در دسترس نیست	تلاش مجدد

قواعد قرارداد: پاسخ باید کمتر از ۲ ثانیه (p95) باشد؛ هر درخواست یک `id_request` یکتا داشته باشد؛ نسخه قواعد در پاسخ ذکر شود؛ و هیچ فیلد شخصی/مالی دقیق بازنگردد. لاگ درخواست فقط شامل متادیتا و شناسه‌های توکن‌شده است.

پیوست ج کاتالوگ کامل قواعد (نمونه اجرایی)

کاتالوگ زیر، نسخه پیشنهادی ۱٫۰ است. هر قاعده پیش از فعال‌سازی، روی داده تاریخی آزمون و توسط کمیته فنی تأیید می‌شود. وزن‌ها در فاز داده کالیبره می‌شوند.

جدول پ-۳ — کاتالوگ قواعد

کد	نام	ورودی	شرط	وزن	وضعیت
01-DUP	تطبیق رخداد بین‌شرکتی	رخداد، خودرو	شباهت ≤ ۰٫۸۵	۲۵	فعال
02-DUP	تکرار خودرو	VIN، تاریخ	دو پرونده در ۷ روز	۲۰	فعال
03-DUP	چندپرونده‌ای	خودرو، ۹۰ روز	≤ ۳ پرونده	۱۵	فعال
01-GAP	شکاف ادعا/پرداخت	مبلغ	≤ ۳۰٪ در ۲ از ۳	۱۲	فعال
02-GAP	بزرگ‌نمایی مکرر	تاریخ، مبلغ	میانگین ≤ ۲۵٪ در ۱۲ ماه	۱۰	فعال
01-STG	صحنه‌سازی تأییدشده	تگ	۲ تگ مستقل	۲۰	فعال
02-STG	تگ مشکوک	تگ	۳ تگ مشکوک	۱۰	فعال
01-NLP	نشانه متنی قوی	متن	۲ برچسب ≤ ۰٫۸	۱۰	فعال
02-NLP	تناقض زمانی	متن، تاریخ	ناسازگاری	۸	فعال
01-REL	تلفن مشترک	گراف	≤ ۳ پرونده	۱۵	فعال
02-REL	کارگاه متمرکز	گراف	≤ ۵۰٪	۱۲	فعال
03-REL	حلقه تبانی	گراف	≤ ۵ پرونده، ۲ شاهد	۲۰	آزمایشی
01-HIS	سابقه پرخسارت	شمارش	≤ ۴ در ۱۲ ماه	۱۰	فعال
01-AMT	مبلغ غیرعادی	توزیع	z ≤ ۳	۸	فعال
01-ID	ناسازگاری هویت	VIN/پلاک	تطبیق ناموفق	۱۵	فعال
01-DOC	سند مشکوک	فراداده	تکرار/ناسازگاری	۰	هشدار
01-TIME	فاصله بیمه‌نامه/حادثه	تاریخ	< ۳۰ روز	۸	فعال
01-EXP	تمرکز کارشناس	گراف	≤ ۴۰٪ پرونده‌ها	۱۰	آزمایشی

حاکمیت قواعد: هر قاعده یک مالک، تاریخ بازبینی و آزمون پس‌رو دارد. تغییر وزن یا شرط، نسخه جدید می‌سازد. قواعد «آزمایشی» فقط در پایلوت سایه اجرا می‌شوند و امتیاز آن‌ها به تصمیم نهایی اضافه نمی‌شود.

پیوست د چک‌لیست داده و اتصال

جدول پ-۴ — چک‌لیست اتصال

#	مورد	وضعیت	مالک
۱	توافق‌نامه تبادل داده امضا شده	☐	حقوقی
۲	نقشه فیلدهای منبع به مدل کانونیکال	☐	داده
۳	نمونه داده بی‌نام/شبه‌واقعی تحویل شده	☐	بیمه‌گر
۴	گزارش کیفیت داده اولیه	☐	داده
۵	دسترسی به محیط sandbox	☐	عملیات
۶	کلید/گواهی mTLS صادر شده	☐	امنیت
۷	اتصال batch فعال	☐	فنی
۸	اتصال API فعال	☐	فنی
۹	آزمون یکپارچه موفق	☐	QA
۱۰	کارشناسان UAT	☐	بیمه‌گر
۱۱	DPIA تأیید شده	☐	DPO
۱۲	آزمون نفوذ انجام شده	☐	امنیت
۱۳	برنامه پاسخ به حادثه	☐	امنیت
۱۴	آموزش کاربران	☐	محصول
۱۵	گزارش نهایی پایلوت	☐	PMO

پیوست ه فرم ارزیابی پایلوت

جدول پ-۵ — فرم ارزیابی

شاخص	مقدار هدف	مقدار واقعی	وضعیت	توضیح
پوشش داده	≥ ۹۰٪		☐	
Precision@50	≥ ۷۵٪		☐	
Recall	≥ ۵۵٪		☐	
مثبت کاذب	≤ ۲۵٪		☐	
زمان پاسخ p95	< ۲s		☐	
دسترس‌پذیری	≥ ۹۹٫۵٪		☐	
کاهش زمان بررسی	≥ ۳۰٪		☐	
رضایت کارشناس	≥ ۴/۵		☐	
ارزش جلوگیری‌شده	≥ ۱۰ میلیارد		☐	

امضاها: نماینده بیمه‌گر: _____ | نماینده ارائه‌دهنده: _____ | ناظر: _____ | تاریخ: _____

پیوست و فرض‌ها و اطلاعات مورد نیاز از کارفرما

برای نهایی‌سازی پیشنهاد، موارد زیر لازم است. بدون این اطلاعات، بخشی از اعداد و طراحی به‌صورت فرض باقی می‌ماند.

جدول پ-۶ — اطلاعات مورد نیاز

#	اطلاعات	چرا لازم است	مسئول	مهلت
۱	نام حقوقی و شناسه شرکت/تیم	شناسنامه پیشنهاد	تیم	پیش از ارسال
۲	CV و تخصیص اعضای تیم	ارزیابی توان تیم	تیم	پیش از ارسال
۳	سه مرجع پروژه مرتبط	اثبات سابقه	تیم	پیش از ارسال
۴	نام دو بیمه‌گر پایلوت	طرح اجرایی	کارفرما	هفته ۱
۵	دیکشنری داده و نمونه	طراحی مدل	بیمه‌گر	هفته ۲
۶	برچسب تقلب و نتایج بررسی	اعتبارسنجی	بیمه‌گر	هفته ۳
۷	نظر حقوقی / قرارداد نمونه	مبنای پردازش	حقوقی	هفته ۲
۸	الزامات امنیتی و اقامت داده	استقرار	امنیت	هفته ۲
۹	تأیید بازه قیمت	مدل مالی	مدیریت	هفته ۲
۱۰	لوگو و هویت بصری	صفحه عنوان	تیم	پیش از ارسال
۱۱	نوع محصول (بدنه/ثالث)	دامنه	کارفرما	هفته ۱
۱۲	حجم تقریبی پرونده سالانه	ظرفیت‌سنجی	بیمه‌گر	هفته ۲

پیوست ز مراجع، استانداردها و چارچوب‌ها

فهرست زیر برای مطالعه و انتخاب چارچوب است؛ ادعای انطباق کامل با هیچ‌کدام بدون ارزیابی مستقل وجود ندارد.

- ISO/IEC 27001 — سیستم مدیریت امنیت اطلاعات
- ISO/IEC 27701 — مدیریت حریم خصوصی و حفاظت داده
- NIST Cybersecurity Framework — چارچوب امنیت سایبری
- OWASP ASVS — استاندارد تأیید امنیت برنامه‌های کاربردی
- OpenAPI Specification 3.0 — قرارداد API
- OpenID Connect / OAuth 2.0 — احراز هویت و مجوز
- HMAC (RFC 2104) — تابع درهم‌سازی کلیددار
- Anonymity-k و Differential Privacy — چارچوب‌های حریم خصوصی
- قانون بیمه و آیین‌نامه‌های بیمه مرکزی جمهوری اسلامی ایران
- قانون تجارت الکترونیک و قانون جرایم رایانه‌ای
- الزامات افتا و مصوبات شورای عالی فضای مجازی
- استانداردهای داده و کدگذاری صنعت بیمه (در صورت وجود)

پایان سند

این پیشنهاد یک نقطه شروع برای گفت‌وگوی فنی و قراردادی است. آماده‌ایم جزئیات را در جلسه فنی، کارگاه داده و پایلوت مشترک نهایی کنیم.